

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	60	100.0	10	21	AAV52143	Mouse EGF derived
2	60	100.0	53	5	AAP40315	Epidermal growth f
3	60	100.0	53	10	AA9P1658	Amino acid sequenc
4	60	100.0	53	11	AAR08007	Modified murine ep
5	60	100.0	53	16	AAR67275	Synthetic human/mo
6	60	100.0	53	18	AAW34466	Chimeric epidermal
7	60	100.0	53	18	AAW34467	Chimeric epidermal
8	60	100.0	53	19	AAW50134	Mouse epidermal gr
9	60	100.0	53	19	AAW50139	Epidermal growth f
10	60	100.0	53	19	AAW50140	Epidermal growth f
11	60	100.0	53	22	AAB37612	Human EGF. Homo s

DR WPI; 2000-013229/01.  
XX New peptide derived from murine epidermal growth factor (mEGF)  
XX  
XX Claim 4; Page 28; 35pp; English.  
XX  
CC This is a peptide derived from mouse epidermal growth factor (EGF)  
CC residues 33-42. This peptide is used in the invention to prepare a  
CC composition to target laminin receptors. EGF derived peptides inhibit  
CC blood vessel formation through their antagonism of the high affinity 67kD  
CC laminin receptor found on endothelial cells. The peptide is modified  
CC from the natural sequence to prevent protease attack. The peptide is used  
CC in the preparation of a medicament for binding to laminin receptors as an  
CC (ant)agonist. The medicament is also useful for healing endothelial cell  
CC wounds and treating angiogenic diseases, especially retinopathy of  
CC immaturity. Other diseases treated include metastatic cancer,  
CC Candida spp. infection, and parasitic infestations like leishmania and  
CC trichomonas vaginalis. The peptide are anti-angiogenic in human models.  
CC The peptides also inhibit both laminin- and EGF-stimulated angiogenesis,  
CC and prevent tumour cell attachment to basement membranes.  
XX  
XX Sequence 10 AA;  
SQ

Query Match 100.0%; Score 60; DB 21; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.0046;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10  
| | | | | | | |  
Db 1 cvigysgdc 10

RESULT 2  
ID AAP40315  
XX AAP40315 standard; protein; 53 AA.  
XX AAP40315;  
XX  
XX 22-JUL-1992 (first entry)  
XX  
XX Epidermal growth factor tripentacontapeptide.  
XX EGF.  
XX Synthetic.  
XX

FH Key Location/Qualifiers  
FT Disulfide-bond 6..20  
FT Disulfide-bond 14..31  
XX Disulfide-bond 33..42  
XX JP59027858-A.  
XX 14-FEB-1984.  
XX  
XX 05-AUG-1982; 82JP-0137128.  
XX  
XX 05-AUG-1982; 82JP-0137128.  
XX  
XX (NNSH ) NIPPON SHINYAKU KK.  
XX  
XX WPI; 1984-072465/12.  
XX

Synthesis of epidermal growth factor polypeptide - by condensation  
PT of protected smaller peptide sequences, de:protection then oxidn. to  
PT cyclise.  
XX

PS Claim1; Page 1; 8pp; Japanese.  
XX The amino acid sequence is that of an epidermal growth factor  
CC tripentacontapeptide which is synthesised by condensation of  
CC protected smaller peptide sequences. This method produces the

CC peptide smoothly, with high purity and yield.  
XX  
XX Sequence 53 AA;  
SQ

Query Match 100.0%; Score 60; DB 5; Length 53;  
Best Local Similarity 100.0%; Pred. No. 0.02;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10  
| | | | | | | |  
Db 33 cvigysgdc 42

RESULT 3  
AAP91658  
ID AAP91658 standard; protein; 53 AA.  
XX AAP91658;  
XX  
XX 29-JUN-1990 (first entry)  
XX  
XX Amino acid sequence for naturally occurring epidermal growth factor  
XX (EGF).  
XX Epidermal growth factor; angiogenesis; synthetic peptide.  
XX  
XX W08901489-A.  
XX  
XX 23-FEB-1989.  
XX  
XX 10-AUG-1988; 88WO-AU00300.  
XX  
XX 10-AUG-1987; 87AU-0003629.  
XX  
XX (CSIR ) COMMONWEALTH SCIENT ORG.  
XX  
XX McAuslan BR;  
XX  
XX WPI; 1989-068852/09.  
XX  
XX Synthetic peptide active in stimulating angiogenesis -  
XX has sequences corresponding to amino acid sequences occurring in  
XX epidermal growth factor.  
XX Fig 1; 1/1; lipp; English.

XX The inventors claim synthetic peptides which correspond to sequences  
XX occurring in EGF, but excluding EGF. The synthetic peptides  
XX correspond to active sequences from EGF fragments: 3-14, 12-15,  
XX 3-10, 29-37 and 33-37 (Claims 2-7 and AAP91659-p91664 resp.). The  
XX peptides are angiogenic. Their relative shortness means that they pose  
XX fewer synthesis problems than the entire EGF molecule.  
XX  
XX Sequence 53 AA;  
SQ

Query Match 100.0%; Score 60; DB 10; Length 53;  
Best Local Similarity 100.0%; Pred. No. 0.02;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10  
| | | | | | | |  
Db 33 cvigysgdc 42

RESULT 4  
AAR08007  
ID AAR08007 standard; protein; 53 AA.  
XX AAR08007;  
XX  
XX 25-FEB-1991 (first entry)  
XX

XX DE Modified murine epidermal growth factor.  
 XX KW Modified murine epidermal growth factor; stability; storage;  
 XX KW epithelial wounds; gastric acid secretion.  
 XX FT Key Location/Qualifiers  
 XX FT Misc-difference 11..11  
 XX FT /label= E, N, Q, A, K  
 XX PN W09013570-A.  
 XX PD 15-NOV-1990.  
 XX PF 09-MAY-1990; 90WO-US02600.  
 XX PR 12-MAY-1989; 89US-0351773.  
 XX PA (CHIR-) CHIRON CORP.  
 XX PI Nasclmento CG, Medina-Selby A;  
 XX DR WPI; 1990-361427/48.  
 XX PT Human epidermal growth factor - is substituted at position 11 for  
 XX PT greater stability and improved storage life.  
 XX PS Claim 9; Page 25; 32pp; English.  
 XX CC The human rEGF is used to treat oversecretion of gastric acid or an  
 XX CC epithelial wound. EGF is modified to increase its chemical  
 XX CC stability. Its storage life is improved without diminishing its  
 XX CC biological activity. The proteins may be prepared by traditional  
 XX CC chemical or recombinant means.  
 XX CC See also AAR08004.  
 XX SQ Sequence 53 AA;  
 Query Match 100.0%; Score 60; DB 11; Length 53;  
 Best Local Similarity 100.0%; Pred. No. 0.02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CVIGYSGDRC 10  
 Db 33 cvigysgdc 42  
 RESULT 5  
 AAR67275  
 ID AAR67275 standard; peptide; 53 AA.  
 AC AAR67275;  
 XX 01-AUG-1995 (first entry)  
 DT Synthetic human/mouse EGF.  
 DE  
 XX Cell growth factor; viscoelastic solution; fibroblastic growth factor;  
 KW FGF; epidermal growth factor; EGF; buffered solution; lubrication;  
 KW carboxypropyl methylcellulose; cellulose gum; dextran; dextran sulphate;  
 KW chondroitin sulphate; sodium hyaluronate; osmolality; mitogenic;  
 KW wound healing; cell protection; cell coating; surgery; tissue space;  
 KW hydroxypropyl methylcellulose; manipulation.  
 XX Synthetic.  
 OS  
 XX Key Location/Qualifiers  
 XX FT Misc-difference 37  
 XX FT /note= "Given in the specification as Try"  
 XX PN US5366964-A.  
 XX

PD 22-NOV-1994.  
 XX 15-DEC-1988; 88US-0284533.  
 XX PR 15-DEC-1988; 88US-0284533.  
 XX PR 13-NOV-1989; 89US-0434305.  
 XX (LIND/) LINDSTROM R L.  
 XX PA (SKEL/) SKELNIK D.  
 XX PI Lindstrom RL, Skelnik D;  
 XX DR WPI; 1995-005865/01.  
 XX PT Viscoelastic soln. contg. hydroxypropyl methylcellulose, sodium  
 XX PT hyaluronate, chondroitin sulphate and growth factors - used as a  
 XX PT surgical soln. to promote wound healing, e.g. of corneal cells  
 XX PS Disclosure; Column 5; 8pp; English.  
 XX CC The sequences given in AAR67273-76 are cell growth factors which may  
 XX CC be used in the viscoelastic solution of the invention. The peptides  
 XX CC are derived from fibroblastic growth factor (EGF) and epidermal  
 XX CC growth factor (EGF). The viscoelastic solution also comprises a  
 XX CC buffered solution which is pref. a buffered balanced salt solution,  
 XX CC at least 1 of hydroxy- or carboxypropyl methylcellulose, a cellulose  
 XX CC gum, dextran or dextran sulphate, chondroitin sulphate, and sodium  
 XX CC hyaluronate. It has a pH of 6-8 and an osmolality of 200-400 mOsmol/L.  
 XX CC The growth factors are mitogenic in vitro for a wide range of tissues  
 XX CC and the viscoelastic soln. may be used as a surgical soln. which is  
 XX CC in direct contact with cells undergoing wound healing. It also provides  
 XX CC a cell protection and cell coating during surgery. The soln. provides  
 XX CC maintenance of tissue space, hydroxypropyl methylcellulose and  
 XX CC chondroitin sulphate lubricate the tissue, while sodium hyaluronate  
 XX CC provides tissue manipulation.  
 XX SQ Sequence 53 AA;  
 Query Match 100.0%; Score 60; DB 16; Length 53;  
 Best Local Similarity 100.0%; Pred. No. 0.02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CVIGYSGDRC 10  
 Db 33 cvigysgdc 42  
 RESULT 6  
 AAW34466  
 ID AAW34466 standard; Protein; 53 AA.  
 XX AC AAW34466;  
 XX 08-MAY-1998 (first entry)  
 DT Chimeric epidermal growth factor protein AC.  
 DE  
 XX Chimeric epidermal growth factor; EGF; chimeric; human; mouse; recombinant.  
 KW Chimeric - Mus sp.  
 XX Chimeric - Homo sapiens.  
 OS  
 XX Key Location/Qualifiers  
 XX FT Region 1..10  
 XX FT /note= "region A from mouse EGF"  
 XX FT Region 11..33  
 XX FT /note= "region B from human EGF"  
 XX FT Region 34..47  
 XX FT /note= "region C from mouse EGF"  
 XX FT Region 48..53  
 XX FT /note= "region D from human EGF"  
 XX

PN AU9717759-A.  
 PD 30-OCT-1997.  
 XX  
 XX  
 PF 08-APR-1997; 97AU-0017759.  
 XX  
 PR 23-APR-1996; 96JP-0123970.  
 XX  
 XX (HGCT ) HIGETA SHOYU KK.  
 PA  
 XX Tagami H, Tanaka A;  
 PI WPI; 1997-550187/51.  
 XX N-PSDB; AAT99941.  
 DR  
 DR Chimeric epidermal growth factor proteins - and DNA molecules for  
 XX their recombinant production  
 XX Claim 5; Page 26; 40pp; English.  
 XX This is a chimeric epidermal growth factor (EGF) protein AC. This  
 CC chimeric EGF protein contains sequences derived from human and mouse  
 CC EGF sequences. The chimeric EGF proteins can be recombinantly produced  
 CC by culturing a microorganism of the genus Bacillus transformed by an  
 CC expression vector containing the encoding DNA molecules. The recombinant  
 CC chimeric EGF proteins exhibit a wide variety of physiological activities  
 CC similar to those of naturally occurring EGF.  
 XX Sequence 53 AA;  
 SQ

Query Match 100.0%; Score 60; DB 18; Length 53;  
 Best Local Similarity 100.0%; Pred. No. 0.02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 CVIGYSGDRC 10  
 Db 33 cvigysgdc 42

RESULT 7  
 AAW34467  
 ID AAW34467 standard; Protein; 53 AA.  
 AC AAW34467;  
 XX  
 DT 08-MAY-1998 (first entry)  
 XX  
 DE Chimeric epidermal growth factor protein ABC.  
 XX  
 KW Epidermal growth factor; EGF; chimeric; human; mouse; recombinant.  
 XX  
 OS Chimeric - Mus sp.  
 XX Chimeric - Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 1..10  
 FT /note= "region A from mouse EGF"  
 FT Region 11..33  
 FT /note= "region B from mouse EGF"  
 FT Region 34..47  
 FT /note= "region C from mouse EGF"  
 FT Region 48..53  
 FT /note= "region D from human EGF"  
 XX  
 PN AU9717759-A.  
 XX  
 PD 30-OCT-1997.  
 XX  
 PF 08-APR-1997; 97AU-0017759.  
 XX  
 PR 23-APR-1996; 96JP-0123970.  
 XX

PA (HGCT ) HIGETA SHOYU KK.  
 XX  
 PI Tagami H, Tanaka A;  
 XX  
 DR WPI; 1997-550187/51.  
 DR N-PSDB; AAT99942.  
 XX  
 XX Chimeric epidermal growth factor proteins - and DNA molecules for  
 PT their recombinant production  
 PT Claim 6; Page 26; 40pp; English.  
 XX  
 XX This is a chimeric epidermal growth factor (EGF) protein ABC. This  
 CC chimeric EGF protein contains sequences derived from human and mouse  
 CC EGF sequences. The chimeric EGF proteins can be recombinantly produced  
 CC by culturing a microorganism of the genus Bacillus transformed by an  
 CC expression vector containing the encoding DNA molecules. The recombinant  
 CC chimeric EGF proteins exhibit a wide variety of physiological activities  
 CC similar to those of naturally occurring EGF.  
 XX Sequence 53 AA;  
 SQ

Query Match 100.0%; Score 60; DB 18; Length 53;  
 Best Local Similarity 100.0%; Pred. No. 0.02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 CVIGYSGDRC 10  
 Db 33 cvigysgdc 42

RESULT 8  
 AAW50134  
 ID AAW50134 standard; protein; 53 AA.  
 XX  
 AC AAW50134;  
 XX  
 DT 08-JUL-1998 (first entry)  
 XX  
 DE Mouse epidermal growth factor.  
 XX  
 KW Mouse; epidermal growth factor; EGF; isolation; recombinant;  
 XX ion exchange chromatography; Bacillus brevis.  
 XX  
 OS Mus sp.  
 XX  
 PN AU9728698-A.  
 XX  
 PD 05-FEB-1998.  
 XX  
 PF 17-JUL-1997; 97AU-0028698.  
 XX  
 PR 01-AUG-1996; 96JP-0218109.  
 XX  
 XX (HGCT ) HIGETA SHOYU KK.  
 PA  
 XX Miyauchi A, Nemoto A, Warren B;  
 PI WPI; 1998-131057/13.  
 DR  
 DR Isolation of recombinant epidermal growth factor from whole broth -  
 PT by fluidised-bed ion-exchange chromatography  
 PT Claim 6; Page 26; 38pp; English.  
 XX  
 XX The present sequence represents mouse epidermal growth factor (EGF)  
 CC from the present invention. The present invention describes a method for  
 CC the isolation of recombinant epidermal growth factor (EGF) from whole  
 CC broth. The method comprises: (a) passing a culture containing the  
 CC protein upwards through a column containing a fluidised bed of ion  
 CC exchanger, washing the ion exchanger to remove non-adsorbed material,  
 CC allowing the ion exchanger to settle, and eluting adsorbed protein by



CC passing an eluant downwards through the column; (b) filtering the eluate  
 CC through a membrane with a molecular weight cutoff of 10000, and (c)  
 CC concentrating the filtrate on a membrane with a molecular weight cutoff  
 CC of 5000. The method is for isolating recombinant human, mouse, pig or  
 CC rat EGF, or an EGF analogue with one of four defined amino acid sequences  
 CC (given in the specification), from *Bacillus brevis* cultures. The EGF  
 CC proteins can be recovered in high yield and high purity without the need  
 CC to pretreat the culture broth to remove the cells, e.g. by costly  
 CC centrifugation or membrane filtration.

XX SQ Sequence 53 AA;

Query Match 100.0%; Score 60; DB 19; Length 53;  
 Best Local Similarity 100.0%; Pred. No. 0.02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10  
 DB 33 cvigysgsrc 42  
 |||||

RESULT 9

AAW50139  
 ID AAW50139 standard; protein; 53 AA.

XX AC AAW50139;

XX DT 08-JUL-1998 (first entry)

XX DE Epidermal growth factor analogue 3.

XX KW Epidermal growth factor; EGF; isolation; recombinant;  
 XX ion exchange chromatography; *Bacillus brevis*.

XX OS Unidentified.

XX PN AU9728698-A.

XX PD 05-FEB-1998.

XX PF 17-JUL-1997; 97AU-0028698.

XX PR 01-AUG-1996; 96JP-0218109.

XX PA (HGET ) HIGETA SHOYU KK.

XX PI Miyauchi A, Nemoto A, Warren B;

XX DR WPI; 1998-131057/13.

XX PT Isolation of recombinant epidermal growth factor from whole broth -  
 XX by fluidised-bed ion-exchange chromatography

XX PS Claim 6; Page 28; 38pp; English.

XX CC The present sequence represents a protein which has epidermal growth  
 CC factor (EGF) like activity, from the present invention. The present  
 CC invention describes a method for the isolation of recombinant epidermal  
 CC growth factor (EGF) from whole broth. The method comprises: (a) passing  
 CC a culture containing the protein upwards through a column containing a  
 CC fluidised bed of ion exchanger, washing the ion exchanger to remove non-  
 CC adsorbed material, allowing the ion exchanger to settle, and eluting  
 CC adsorbed protein by passing an eluant downwards through the column; (b)  
 CC filtering the eluate through a membrane with a molecular weight cutoff  
 CC of 10000, and (c) concentrating the filtrate on a membrane with a  
 CC molecular weight cutoff of 5000. The method is for isolating recombinant  
 CC human, mouse, pig or rat EGF, or an EGF analogue with one of four  
 CC defined amino acid sequences (given in the specification), from *Bacillus*  
 CC *brevis* cultures. The EGF proteins can be recovered in high yield and  
 CC high purity without the need to pretreat the culture broth to remove the  
 CC cells, e.g. by costly centrifugation or membrane filtration.

XX

SQ Sequence 53 AA;

Query Match 100.0%; Score 60; DB 19; Length 53;  
 Best Local Similarity 100.0%; Pred. No. 0.02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10  
 DB 33 cvigysgsrc 42  
 |||||

RESULT 10

AAW50140  
 ID AAW50140 standard; protein; 53 AA.

XX AC AAW50140;

XX DT 08-JUL-1998 (first entry)

XX DE Epidermal growth factor analogue 4.

XX KW Epidermal growth factor; EGF; isolation; recombinant;  
 XX ion exchange chromatography; *Bacillus brevis*.

XX OS Unidentified.

XX PN AU9728698-A.

XX PD 05-FEB-1998.

XX PF 17-JUL-1997; 97AU-0028698.

XX PR 01-AUG-1996; 96JP-0218109.

XX PA (HGET ) HIGETA SHOYU KK.

XX PI Miyauchi A, Nemoto A, Warren B;

XX DR WPI; 1998-131057/13.

XX PT Isolation of recombinant epidermal growth factor from whole broth -  
 XX by fluidised-bed ion-exchange chromatography

XX PS Claim 6; Page 29; 38pp; English.

XX CC The present sequence represents a protein which has epidermal growth  
 CC factor (EGF) like activity, from the present invention. The present  
 CC invention describes a method for the isolation of recombinant epidermal  
 CC growth factor (EGF) from whole broth. The method comprises: (a) passing  
 CC a culture containing the protein upwards through a column containing a  
 CC fluidised bed of ion exchanger, washing the ion exchanger to remove non-  
 CC adsorbed material, allowing the ion exchanger to settle, and eluting  
 CC adsorbed protein by passing an eluant downwards through the column; (b)  
 CC filtering the eluate through a membrane with a molecular weight cutoff  
 CC of 10000, and (c) concentrating the filtrate on a membrane with a  
 CC molecular weight cutoff of 5000. The method is for isolating recombinant  
 CC human, mouse, pig or rat EGF, or an EGF analogue with one of four  
 CC defined amino acid sequences (given in the specification), from *Bacillus*  
 CC *brevis* cultures. The EGF proteins can be recovered in high yield and  
 CC high purity without the need to pretreat the culture broth to remove the  
 CC cells, e.g. by costly centrifugation or membrane filtration.

SQ Sequence 53 AA;

Query Match 100.0%; Score 60; DB 19; Length 53;  
 Best Local Similarity 100.0%; Pred. No. 0.02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10  
 DB 33 cvigysgsrc 42  
 |||||

```

RESULT 11
ID AAB37612 standard; protein; 53 AA.
XX
AC AAB37612;
XX
DT 27-FEB-2001 (first entry)
XX
DE Human EGF.
XX
KW Merozoite surface protein; protazoacide; vaccine; malaria; human; EGF;
KW epidermal growth factor.
XX
OS Homo sapiens.
XX
PN W0200063245-A2.
XX
PD 26-OCT-2000.
XX
PF 20-APR-2000; 2000WO-CB01558.
XX
PR 20-APR-1999; 99GB-0009072.
XX
PR 13-MAY-1999; 99US-0311817.
XX
PR 25-MAY-1999; 99CA-2271451.
XX
PA (MEDI-) MEDICAL RES COUNCIL.
XX
PI Holder A, Birdsell B, Feeney J, Morgan W, Syed S, Uthaipibull C;
XX
DR WPI; 2001-015762/02.
XX
PT Novel variants of the C-terminal fragment of Plasmodium merozoite
PT surface protein-1, useful as vaccines for treating or preventing
XX malaria -
XX
FS Disclosure; Fig 1; 126pp; English.
XX
CC The present invention relates to non-natural variants of a C-terminal
CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC non-natural variants have reduced affinity for at least 1 antibody
CC capable of blocking a second antibody that inhibits the proteolytic
CC cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least
CC one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC MSP-1.4.2, compared to natural MSP-1.9. The non-natural variants of
CC the present invention are useful for immunising a mammal against malaria,
CC and can be used to treat malaria. The present sequence is human
CC epidermal growth factor (EGF). This sequence was used in a sequence
CC homology comparison with the wild-type MSP-1 protein from P. falciparum
CC (see AAB37608), which was used to generate the variants of the present
XX invention.
XX
SQ Sequence 53 AA;

Query Match 100.0%; Score 60; DB 22; Length 53;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVIGYSGDRC 10
Db 33 cvigysgdc 42
|||||

RESULT 12
ID AAP60791 standard; protein; 56 AA.
XX
AC AAP60791;
XX
DT 23-JUN-1991 (first entry)
XX

```

```

DE Epidermal growth factor.
XX
KW Epidermal growth factor analogue.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT Protein 4..56
XX /label= EGF protein
XX
PN GB2172890-A.
XX
PD 01-OCT-1986.
XX
PF 24-MAR-1986; 86GB-0007203.
XX
PR 25-MAR-1985; 85GB-0007666.
XX
PA (WELL ) WELLCOME FOUNDATION LTD.
XX
PI Allen G, Winther MD;
XX
DR WPI; 1986-259985/40.
DR N-PSDB; AAN60708.
XX
PT DNA sequence encoding epidermal growth factor - with carrier
PT protein linked via lysine residue.
XX
PS Fig 1; Page1; 12pp; English.
XX
CC The sequence encodes a mouse EGF fusion protein containing a Lys
CC residue at the N-terminal. This EGF analogue is useful for
CC depilating an animal, especially for defleecing sheep.
XX
SQ Sequence 56 AA;

Query Match 100.0%; Score 60; DB 7; Length 56;
Best Local Similarity 100.0%; Pred. No. 0.021;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CVIGYSGDRC 10
Db 36 cvigysgdc 45
|||||

RESULT 13
ID AAW68455 standard; protein; 117 AA.
XX
AC AAW68455;
XX
DT 21-DEC-1998 (first entry)
XX
DE Recombinant human epidermal growth factor precursor.
XX
KW Phagemid pT72; integration; transposon; Bordetella; repeat sequence;
KW chromosome; bacteriophage; T7; precursor; protease; proinsulin;
KW epidermal growth factor.
XX
OS Synthetic.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 65..117
XX /note= "mature epidermal growth factor portion"
XX
PN W09804731-A1.
XX
PD 05-FEB-1998.
XX
PF 25-JUL-1996; 96WO-RU00198.
XX

```

PR 25-JUL-1996; 96WO-RU00198.

XX (NIKA=) NIKA-UNIVERSAL FIRM.

XX Fedchenko VI, Guriev SO, Maslikova AN, Nechaev VN;

PI Sivov IG;

XX WPI; 1998-130717/12.

XX producing biologically active peptide(s) - comprises separate

PT formation of precursor and protease, useful for, e.g. production of

PT epidermal growth factors

XX Example 12; Page 221; 306pp; Russian.

XX This sequence represents a recombinant human epidermal growth factor

XX precursor protein. The coding sequence for this protein is used in a

CC method for producing biologically active peptides based on binary

CC technology, comprising separate formation of a precursor of the active

CC peptide and a protease for processing the precursor. The method uses

CC the phagemid pT72 (AAV60642) which is derived by integration of a

CC transposon based on a Bordetella repeat sequence into the chromosome of

CC T7 phage, as an expression vector. The methods and the products may be

CC used for the production of proinsulin and epidermal growth factors.

XX SQ Sequence 117 AA;

XX Query Match 100.0%; Score 60; DB 19; Length 117;

XX Best Local Similarity 100.0%; Pred. No. 0.04; Indels 0; Gaps 0;

XX Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10

DB 97 cvigysgdc 106

RESULT 14

AAAP50296

ID AAP50296 standard; Protein; 1080 AA.

XX AAP50296;

AC AAP50296;

XX 03-DEC-1991 (first entry)

XX Mouse epidermal growth factor (EGF) precursor protein.

XX Epidermal growth factor precursor.

XX Mus musculus.

XX W08500369-A.

XX 31-JAN-1985.

XX 02-JUL-1984; 84WO-US01050.

XX 05-JUL-1983; 83US-0511372.

XX (CHIR-) CHIRON CORP.

XX Graeme BI;

XX WPI; 1985-038094/06.

XX N-PSDB; AAN50342.

XX New human DNA sequence encoding epidermal growth factor - useful

XX for prodn. of EGF and related poly:peptide(s).

XX Disclosure; Page 11-12a; 21pp; English.

XX The sequence encodes mouse EGF precursor protein, which is a potent

XX mitogen for a variety of cells, and a potent inhibitor of gastric

CC

CC acid secretion.

XX Sequence 1080 AA;

XX Query Match 88.3%; Score 53; DB 6; Length 1080;

XX Best Local Similarity 90.0%; Pred. No. 3.1;

XX Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10

DB 1009 cvigysgdc 1018

RESULT 15

AAW34460

ID AAW34460 standard; peptide; 14 AA.

XX AAW34460;

AC AAW34460;

XX 08-MAY-1998 (first entry)

XX Chimeric epidermal growth factor protein partial peptide sequence 3.

XX Epidermal growth factor; EGF; chimeric; human; mouse; recombinant.

XX Mus sp.

XX Homo sapiens.

XX AU9717759-A.

XX 30-OCT-1997.

XX 08-APR-1997; 97AU-0017759.

XX 23-APR-1996; 96JP-0123970.

XX (HGET ) HIGETA SHOYU KK.

XX Tagami H, Tanaka A;

XX WPI; 1997-550187/51.

XX Chimeric epidermal growth factor proteins - and DNA molecules for

XX their recombinant production

XX Claim 2; Page 23; 40pp; English.

XX This peptide sequence corresponds to the C region of a chimeric epidermal

XX growth factor (EGF) protein. This chimeric EGF protein contains sequences

XX derived from human and mouse EGF sequences. These chimeric proteins can

XX be recombinantly produced by culturing a microorganism of the genus

XX Bacillus transformed by an expression vector containing the encoding DNA

XX molecules. The recombinant chimeric EGF proteins exhibit a wide variety

XX of physiological activities similar to those of naturally occurring EGF.

XX SQ Sequence 14 AA;

XX Query Match 85.0%; Score 51; DB 18; Length 14;

XX Best Local Similarity 100.0%; Pred. No. 0.14;

XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 VIGYSGDRC 10

DB 1 vigsygdrc 9

Search completed: August 15, 2001, 10:54:54

Job time: 207 sec

11

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2001, 10:51:27 ; Search time 22.89 Seconds  
(without alignments)  
33.279 Million cell updates/sec

Title: US-09-673-785A-2

Perfect score: 60

Sequence: 1 CVIGYSGDRC 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	60	100.0	1217	1	EGMSMG	epidermal growth f
2	55	91.7	2139	2	A35672	crumbs protein - f
3	53	88.3	1643	2	T14274	versican precursor
4	53	88.3	3381	2	T42389	versican precursor
5	50	83.3	57	2	PC4415	ErbB kinase activa
6	50	83.3	372	2	T29359	hypothetical prote
7	50	83.3	1133	1	EGRT	epidermal growth f
8	49	81.7	102	2	B55885	chondroitin sulfat
9	49	81.7	862	2	S43922	versican - pig-tai
10	49	81.7	1207	1	EGHU	epidermal growth f
11	49	81.7	2409	1	A60979	versican precursor
12	48	80.0	2352	2	T30201	Notch homolog prot
13	48	80.0	4006	2	T09070	probable tenascin
14	46	76.7	53	2	S17294	epidermal growth f
15	46	76.7	57	2	B69300	hypothetical prote
16	45	75.0	1531	2	T42218	slit-1 protein hom
17	45	75.0	2397	1	A55335	versican precursor
18	44	73.3	3106	1	S53868	laminin alpha-2 ch
19	44	73.3	3672	2	T23433	hypothetical prote
20	44	73.3	3704	2	T37316	probable laminin a
21	43	71.7	264	2	T16271	hypothetical prote
22	43	71.7	447	2	A39321	mucin - rat (fragm
23	43	71.7	1064	2	T27358	hypothetical prote
24	43	71.7	1054	2	A40136	fibropellin Ia - s
25	43	71.7	3097	2	T00021	DN-cadherin - frui
26	42	70.0	832	2	A31246	neurogenic protein
27	42	70.0	833	2	S19087	gene Delta protein
28	42	70.0	880	2	S00670	neurogenic repetit
29	42	70.0	883	2	S49126	brevicain precursor

## ALIGNMENTS

RESULT 1

EGMSMG

epidermal growth factor precursor - mouse

N:Alternate names: urogastrone precursor

C:Species: Mus musculus (house mouse)

C:Date: 30-Nov-1980 #sequence\_revision 11-Aug-1983 #text\_change 19-Jan-2001

C:Accession: A94272; A93304; A92118; A01387

R:Scott, J.; Urdea, M.; Quiroga, M.; Sanchez-Pescador, R.; Fong, N.; Selby, M.; Rutte

Science 221, 236-240, 1983

A:Title: Structure of a mouse submaxillary messenger RNA encoding epidermal growth fa

A:Reference number: A94272; MUID:83223630

A:Accession: A94272

A:Molecule type: mRNA

A:Residues: 1-1217 <SCO>

A:Cross-references: GB:J00380; NID:gl92993; PIDN:AAA37539.1; PID:g309210

R:Gray, A.; Dull, T.; Ullrich, A.

Nature 303, 722-725, 1983

A:Title: Nucleotide sequence of epidermal growth factor cDNA predicts a 128,000-molec

A:Reference number: A93304; MUID:83219309

A:Accession: A93304

A:Molecule type: mRNA

A:Residues: 1-789, 'Y', 791-1047, 'S', 1049-1168 <GRA>

A:Cross-references: GB:J00380

A:Note: the sequence shown by these authors differs from residues 1134-1168 due to an

uence of Scott et al.)

R:Savage Jr., C.R.; Inagami, T.; Cohen, S.

J. Biol. Chem. 247, 7612-7621, 1972

A:Title: The primary structure of epidermal growth factor.

A:Reference number: A92118; MUID:73048516

A:Accession: A92118

A:Molecule type: protein

A:Residues: 977-1029 <SAV>

A:Note: residues 1024-1029 are not required for full biological activity in vivo

R:Savage Jr., C.R.; Hash, J.H.; Cohen, S.

J. Biol. Chem. 248, 7669-7672, 1973

A:Title: Epidermal growth factor. Location of disulfide bonds.

A:Reference number: A92144; MUID:74025498

A:Contents: annotation; disulfide bonds

C:Comment: Epidermal growth factor (EGF) stimulates the proliferation and differentia

gastrointestinal cell proliferation.

C:Comment: EGF is released in the pancreas, small intestine, mammary gland, and (in s

C:Comment: The EGF precursor is found in kidney as a receptor-like membrane-bound pro

C:Comment: The active growth factor from this submaxillary gland protein stimulates t

C:Superfamily: epidermal growth factor precursor; EGF homology; LDL receptor YWTD-con

C:Keywords: duplication; growth factor; tandem repeat; transmembrane protein

F:1-28/Domain: signal sequence #status predicted <Sig>

F:29-1217/Product: epidermal growth factor proprotein, membrane-bound form #status pr

F:29-1038/Domain: extracellular #status predicted <EXT>

F:50-485/Region: EGF precursor long repeat

F:53-92/Domain: LDL receptor YWTD-containing repeat homology <YW01>

F:93-134/Domain: LDL receptor YWTD-containing repeat homology <YW02>

F:135-176/Domain: LDL receptor WYTD-containing repeat homology <YW03>  
 F:177-217/Domain: LDL receptor WYTD-containing repeat homology <YW04>  
 F:218-262/Domain: LDL receptor WYTD-containing repeat homology <YW04>  
 F:263-307/Domain: LDL receptor WYTD-containing repeat homology <YW05>  
 F:324-360/Domain: LDL receptor WYTD-containing repeat homology <YW06>  
 F:366-401/Domain: EGF homology <EG2>  
 F:407-442/Domain: EGF homology <EG2>  
 F:445-482/Domain: EGF homology <EG3>  
 F:486-529/Domain: EGF precursor long repeat  
 F:530-572/Domain: LDL receptor WYTD-containing repeat homology <YW07>  
 F:573-615/Domain: LDL receptor WYTD-containing repeat homology <YW08>  
 F:616-659/Domain: LDL receptor WYTD-containing repeat homology <YW09>  
 F:660-700/Domain: LDL receptor WYTD-containing repeat homology <YW10>  
 F:701-743/Domain: LDL receptor WYTD-containing repeat homology <YW11>  
 F:751-786/Domain: EGF homology <EG5>  
 F:845-875/Domain: EGF homology <EG6>  
 F:881-917/Domain: EGF homology <EG7>  
 F:923-958/Domain: EGF homology <EG8>  
 F:977-1029/Product: epidermal growth factor #status experimental <EGF>  
 F:982-1018/Domain: EGF homology <EG9>  
 F:1039-1063/Domain: transmembrane #status predicted <TM>  
 F:1064-1217/Domain: intracellular #status predicted <INT>  
 F:347-360,366-377,373-386,388-401,407-418,414-427,429-442,445-457,453-467,469-482,751-786  
 tus predicted  
 F:982-996,990-1007,1009-1018/Disulfide bonds: #status experimental

Query Match 100.0%; Score 60; DB 1; Length 1217;  
 Best Local Similarity 100.0%; Pred. No. 0.019;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10  
 |||||  
 Db 1009 CVIGYSGDRC 1018

RESULT 2  
 A35672

Crumbs protein - fruit fly (Drosophila melanogaster)  
 C:Species: Drosophila melanogaster  
 C>Date: 21-Sep-1990 #sequence\_revision 18-Nov-1992 #text\_change 11-Jan-2000  
 C:Accession: A35672  
 R:Teppas, U.; Theres, C.; Knust, E.  
 Cell 61, 787-799, 1990  
 A:Title: crumbs encodes an EGF-like protein expressed on apical membranes of Drosophila  
 A:Reference number: A35672; MUID:90263104  
 A:Accession: A35672  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-2139 <TEP>  
 A:Cross-references: GB:M33753  
 A:Note: the authors translated the codon GGC for residue 1928 as Cys, and TAT for residue  
 C:Genetics:  
 A:Gene: FlyBase:crb  
 A:Cross-references: FlyBase:FBgn000368  
 C:Superfamily: unassigned EGF-related proteins; EGF homology  
 C:Keywords: transmembrane protein  
 F:352-385/Domain: EGF homology <EGX1>  
 F:392-424/Domain: EGF homology <EGF1>  
 F:691-722/Domain: EGF homology <EGF>  
 F:767-799/Domain: EGF homology <EGF3>  
 F:1878-1914/Domain: EGF homology <EGX2>

Query Match 91.7%; Score 55; DB 2; Length 2139;  
 Best Local Similarity 80.0%; Pred. No. 0.23;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10  
 |||||  
 Db 571 CAVIGYSGDRC 580

RESULT 3  
 T14274

versican precursor, splice form V2 - bovine  
 C:Species: Bos primigenius taurus (cattle)  
 C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 05-May-2000  
 C:Accession: T14274  
 R:Schmalfeldt, M.; Dours-Zimmermann, M.T.; Winterhalter, K.H.; Zimmermann, D.R.  
 J. Biol. Chem. 273, 15758-15764, 1998

A:Title: Versican V2 is a major extracellular matrix component of the mature bovine b  
 A:Reference number: Z17954; MUID:98288320  
 A:Accession: T14274  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1643 <SCH>  
 A:Cross-references: EMBL:AF060458; NID:g32533303; PID:g32533304; PIDN:AAC24360.1  
 A:Experimental source: brain  
 C:Keywords: glycoprotein  
 F:1-20/Domain: signal sequence #status predicted <SIG>  
 F:21-1643/Product: versican, splice form V2 #status predicted <MAT>  
 F:57,331,352,817,965,1017,1333,1616,1626/Binding site: carbohydrate (Asn) (covalent)

Query Match 88.3%; Score 53; DB 2; Length 1643;  
 Best Local Similarity 90.0%; Pred. No. 0.39;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10  
 |||||  
 Db 1362 CVPGYSGDRC 1371

RESULT 4  
 T42389

versican precursor, splice form V0 - bovine  
 N:Alternate names: chondroitin sulfate proteoglycan  
 C:Species: Bos primigenius taurus (cattle)  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 05-May-2000  
 C:Accession: T42389  
 R:Schmalfeldt, M.; Dours-Zimmermann, M.T.; Winterhalter, K.H.; Zimmermann, D.R.  
 J. Biol. Chem. 273, 15758-15764, 1998  
 A:Title: Versican V2 is a major extracellular matrix component of the mature bovine  
 A:Reference number: Z17954; MUID:98288320  
 A:Accession: T42389  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-3381 <SCH>

A:Cross-references: EMBL:AF060456; NID:g3253299; PID:g3253300; PIDN:AAC24358.1  
 C:Superfamily: chicken chondroitin sulfate proteoglycan PG-M core protein; C-type lec  
 C:Keywords: chondroitin sulfate proteoglycan; extracellular matrix; glycoprotein  
 F:1-20/Domain: signal sequence #status predicted <SIG>  
 F:21-3381/Product: versican, splice form V0 #status predicted <MAT>  
 F:57,331,352,817,965,1017,1333,1393,1437,1463,1653,1974,2045,2074,2103,2263,2290,2356

Query Match 88.3%; Score 53; DB 2; Length 3381;  
 Best Local Similarity 90.0%; Pred. No. 0.75;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10  
 |||||  
 Db 3100 CVPGYSGDRC 3109

RESULT 5  
 PC4415

ErbB kinase activator beta, brain and thymus - rat (fragment)  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 10-Dec-1997 #sequence\_revision 10-Dec-1997 #text\_change 11-Jan-2000  
 C:Accession: PC4415  
 R:Higashiyama, S.; Horikawa, M.; Yamada, K.; Ichino, N.; Nakano, N.; Nakagawa, T.; Ml  
 J. Biochem. 122, 675-680, 1997  
 A:Title: A novel brain-derived member of the epidermal growth factor family that inte

A:Reference number: JC5700; MUID:98006324  
 A:Accession: PC4415  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-57 <HIG>

A:Cross-references: DDBJ:D89997; NID:g2605633; PIDN:BAA23346.1; PID:g2605634  
 A:Experimental source: PC-12 cell  
 C:Comment: This protein is a member of the epidermal growth factor family. It is functional in the differentiation of MDA-MB-453 cells.  
 C:Superfamily: unassigned EGF-related proteins; EGF homology  
 F:1-25/Domain: EGF homology (fragment) <EGF>

Query Match 83.3%; Score 50; DB 2; Length 57;  
 Best Local Similarity 70.0%; Pred. NO. 0.061;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGYSGDRC 10

Db 16 CPVGTGDC 25

RESULT 6

T29359

hypothetical protein R05G6.9 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T29359

R:Murray, J.; Le, T.T.

submitted to the EMBL Data Library, May 1996

A:Description: The sequence of C. elegans cosmid R05G6.

A:Reference number: Z20612

A:Accession: T29359

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-372 <MUR>

A:Cross-references: EMBL:U58746; PIDN:AAB00626.1; GSPDB:GN00022; CESP:R05G6.9

A:Experimental source: strain Bristol N2; clone R05G6

C:Genetics:

A:Gene: CESP:R05G6.9

A:Map position: 4

A:Introns: 80/1; 161/1; 245/1; 286/1

Query Match 83.3%; Score 50; DB 2; Length 372;

Best Local Similarity 70.0%; Pred. NO. 0.34;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGYSGDRC 10

Db 233 CYLGYSGDKC 242

RESULT 7

EGRT

epidermal growth factor precursor - rat

N:Alternate names: urogastone precursor

C:Species: Rattus norvegicus (Norway rat)

C:Date: 31-Dec-1998 #sequence\_revision 14-Aug-1998 #text\_change 18-Jun-1999

C:Accession: I52995; S05074; S01974; A25425; S18419; S08286

R:Saggi, S.J.; Safirstein, R.; Price, P.M.

DNA Cell Biol. 11, 481-487, 1992

A:Title: Cloning and Sequencing of the Rat Preproepidermal Growth Factor cDNA: Comparison

A:Reference number: I52995; MUID:92398779

A:Accession: I52995

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1133 <RES>

A:Cross-references: EMBL:U04842; NID:g440236; PIDN:AAB60436.1; PID:g440237

R:Simpson, R.

submitted to the EMBL Data Library, August 1988

A:Reference number: S05074

A:Accession: S05074

Query Match

83.3%; Score 50; DB 1; Length 1133;

A:Molecule type: mRNA  
 A:Residues: 'W', 966; 'RWL', 970-1023, 'NW', 1026-1108, 'SGAGVSGGPWFVLE', 1126, 'HQ' <SIM>  
 A:Cross-references: EMBL:X12748  
 R:Dorow, D.S.; Simpson, R.J.  
 Nucleic Acids Res. 16, 9338, 1988  
 A:Title: Cloning and sequence analysis of a cDNA for rat epidermal growth factor.  
 A:Reference number: S01974; MUID:89016634  
 A:Accession: S01974  
 A:Molecule type: mRNA  
 A:Residues: 'W', 966; 'RWL', 970-1023, 'NW', 1026-1108 <DOR>  
 A:Cross-references: EMBL:X12748  
 R:Simpson, R.J.; Smith, J.A.; Moritz, R.L.; O'Hare, M.J.; Rudland, P.S.; Morrison, J.  
 Eur. J. Biochem. 153, 629-637, 1985  
 A:Title: Rat epidermal growth factor: complete amino acid sequence.  
 A:Reference number: A25425; MUID:86081810  
 A:Accession: A25425  
 A:Molecule type: protein  
 A:Residues: 974-1021 <SL2>  
 R:Nishi, N.; Shimizu, C.; Okutani, T.; Kagawa, Y.; Takasuga, H.; Suno, M.; Wada, F.  
 Biochim. Biophys. Acta 1095, 268-275, 1991  
 A:Title: Rat prostatic growth factors: purification and characterization of high and  
 A:Reference number: S18419; MUID:92069070  
 A:Accession: S18419  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 974-1021 <NIS>  
 R:Nexo, E.; Jorgensen, P.E.; Thim, L.; Roepstorff, P.  
 Biochim. Biophys. Acta 1037, 388-393, 1990  
 A:Title: Purification and characterization of a low and a high molecular weight form  
 A:Reference number: S08288; MUID:90181442  
 A:Accession: S08288  
 A:Molecule type: protein  
 A:Residues: 974-1024 <NEX>  
 C:Comment: Epidermal growth factor (EGF) stimulates the proliferation and differentiation  
 gastrointestinal cell proliferation  
 C:Comment: EGF is released in the pancreas, small intestine, mammary gland, and (in  
 C:Comment: The EGF precursor is found in kidney as a receptor-like membrane-bound pro  
 C:Superfamily: epidermal growth factor precursor; EGF homology; LDL receptor YWTD-con  
 C:Keywords: duplication; growth factor; mitogen; tandem repeat; transmembrane protein  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-1133/Product: epidermal growth factor proprotein, membrane-bound form #status pr  
 F:22-1035/Domain: extracellular #status predicted <EXT>  
 F:44-480/Region: EGF precursor long repeat <LRI>  
 F:47-86/Domain: LDL receptor YWTD-containing repeat homology <YW01>  
 F:87-128/Domain: LDL receptor YWTD-containing repeat homology <YW02>  
 F:129-170/Domain: LDL receptor YWTD-containing repeat homology <YW03>  
 F:171-212/Domain: LDL receptor YWTD-containing repeat homology <YW04>  
 F:213-257/Domain: LDL receptor YWTD-containing repeat homology <YW05>  
 F:258-302/Domain: LDL receptor YWTD-containing repeat homology <YW06>  
 F:319-355/Domain: EGF homology #status atypical <EG1>  
 F:361-396/Domain: EGF homology <EG2>  
 F:402-437/Domain: EGF homology <EG3>  
 F:440-477/Domain: EGF homology <EG4>  
 F:482-958/Region: EGF precursor long repeat <LR2>  
 F:485-525/Domain: LDL receptor YWTD-containing repeat homology <YW07>  
 F:526-568/Domain: LDL receptor YWTD-containing repeat homology <YW08>  
 F:569-611/Domain: LDL receptor YWTD-containing repeat homology <YW09>  
 F:612-655/Domain: LDL receptor YWTD-containing repeat homology <YW10>  
 F:656-696/Domain: LDL receptor YWTD-containing repeat homology <YW11>  
 F:697-739/Domain: LDL receptor YWTD-containing repeat homology <YW12>  
 F:747-782/Domain: EGF homology <EG5>  
 F:839-872/Domain: EGF homology <EG6>  
 F:878-914/Domain: EGF homology <EG7>  
 F:920-955/Domain: EGF homology <EG8>  
 F:974-1024/Product: epidermal growth factor #status experimental <MAT>  
 F:979-1015/Domain: EGF homology <EG9>  
 F:1036-1060/Domain: transmembrane #status predicted <TM>  
 F:1061-1133/Domain: intracellular #status predicted <INT>  
 F:342-355,361-372,368-381,383-396,402-413,409-422,424-437,440-452,448-462,464-477,747  
 -1015/Disulfide bonds: #status predicted

Best Local Similarity 80.0%; Pred. No. 0.92;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10  
|||||I:|I|  
Db 1006 CVIGYIGERC 1015

## RESULT 8

B55885

C:Species: Homo sapiens (man)  
C:Date: 19-Oct-1995 #sequence\_revision 19-Oct-1995 #text\_change 16-Jul-1999  
C:Accession: B55885  
R:Zako, M.; Shinomura, T.; Ujita, M.; Ito, K.; Kimata, K.  
J. Biol. Chem. 270, 3914-3918, 1995  
A:Title: Expression of PG-M(V3), an alternatively spliced form of PG-M without a chondroitin sulfate proteoglycan (man)  
A:Reference number: A55885; MUID:95181355  
A:Accession: B55885  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-102 <ZAK>  
A:Cross-references: GB:S75879; GB:D32039  
C:Superfamily: versican; C-type lectin homology; complement factor H repeat homology; EGF  
F:24-55/Domain: EGF homology <EG1>  
F:62-93/Domain: EGF homology <EG2>

## Query Match

81.7%; Score 49; DB 2; Length 102;

Best Local Similarity 80.0%; Pred. No. 0.15;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10  
|| |||||I|  
Db 46 CVPYSGDQC 55

## RESULT 9

S43922

versican - pig-tailed macaque (fragments)  
N:Alternate names: chondroitin sulfate proteoglycan  
C:Species: Macaca nemestrina (pig-tailed macaque)  
C:Date: 06-Dec-1996 #sequence\_revision 06-Dec-1996 #text\_change 16-Jul-1999  
C:Accession: S43922  
R:Yao, L.Y.; Moody, C.; Schoenherr, E.; Wight, T.N.; Sandell, L.J.  
Matrix Biol. 14, 213-225, 1994  
A:Title: Identification of the proteoglycan versican in aorta and smooth muscle cells by  
A:Reference number: S43921; MUID:95005762  
A:Accession: S43922  
A:Molecule type: mRNA  
A:Residues: 1-233;234-525;526-862 <YAO>  
A:Cross-references: EMBL:S72413  
A:Note: 507-Ser was also found  
A:Note: the authors translated the codon GCC for residue 50 as Val, AAG for residue 422  
C:Superfamily: versican; C-type lectin homology; complement factor H repeat homology; EGF  
F:1-37/Domain: chondroitin sulfate proteoglycan; extracellular matrix  
F:58-139/Domain: link protein repeat homology (fragment) <LNK1>  
F:722-753/Domain: link protein repeat homology <LNK2>  
F:760-791/Domain: EGF homology <EG1>  
F:760-791/Domain: EGF homology <EG2>

## Query Match

81.7%; Score 49; DB 2; Length 862;

Best Local Similarity 80.0%; Pred. No. 1.1;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10  
|| |||||I|  
Db 744 CVPYSGDQC 753

## RESULT 10

## EGHU

epidermal growth factor precursor [validated] - human  
N:Alternate names: urogastrone precursor  
C:Species: Homo sapiens (man)  
C:Date: 30-Nov-1980 #sequence\_revision 14-Aug-1998 #text\_change 08-Dec-2000  
C:Accession: A25531; A01388; A33517; A29721; S45282; S45283  
R:Bell, G.I.; Fong, N.W.; Stempien, M.M.; Wormsted, M.A.; Caput, D.; Ku, L.; Urdea, M.  
Nucleic Acids Res. 14, 8427-8446, 1986  
A:Title: Human epidermal growth factor precursor: cDNA sequence, expression in vitro  
A:Reference number: A25531; MUID:87066721  
A:Accession: A25531  
A:Molecule type: mRNA  
A:Residues: 1-1207 <BEL>  
A:Cross-references: EMBL:X04571; NID:g31120; PIDN:CAA28240.1; PID:g31121  
A:Note: 708-Met was also found  
A:Note: intron positions were also determined  
R:Gregory, H.; Preston, B.M.  
Int. J. Pept. Protein Res. 9, 107-118, 1977  
A:Title: The primary structure of human urogastrone.  
A:Reference number: A01388; MUID:77117897  
A:Accession: A01388  
A:Molecule type: protein  
A:Residues: 971-1023 <GRE>  
A:Note: some of the molecules lack Arg-1023  
R:Furuya, M.; Akashi, S.; Hirayama, K.  
Biochem. Biophys. Res. Commun. 163, 1100-1106, 1989  
A:Title: The primary structure of human EGF produced by genetic engineering, studied  
A:Reference number: A33517; MUID:89391964  
A:Accession: A33517  
A:Molecule type: protein  
A:Residues: 971-1023 <FUR>  
R:Tsukumo, K.; Nakamura, H.; Sakamoto, S.  
Biochem. Biophys. Res. Commun. 145, 126-133, 1987  
A:Title: Purification and characterization of high molecular weight human epidermal  
A:Reference number: A29721; MUID:87241488  
A:Accession: A29721  
A:Molecule type: protein  
A:Residues: 829-834,'X',836-839,'X',841-845,'X',847-848 <TSU>  
A:Note: this is the amino-terminal sequence of a high molecular weight form of EGF, 1  
R:Svoboda, M.; Bauhofer, A.; Schwind, P.; Bade, E.; Rasched, I.; Przybylski, M.  
Biochim. Biophys. Acta 1206, 35-41, 1994  
A:Title: Structural characterization and biological activity of recombinant human epi  
A:Reference number: S45282; MUID:94242778  
A:Accession: S45282  
A:Molecule type: protein  
A:Residues: 'M',971-1023 <SVO>  
A:Note: expressed recombinant protein  
A:Accession: S45283  
A:Molecule type: protein  
A:Residues: 'MKYP',970-1023 <SV2>  
A:Note: expressed recombinant protein  
C:Comment: Epidermal growth factor (EGF) stimulates the proliferation and differentia  
gastrointestinal cell proliferation.  
C:Comment: EGF is released in the pancreas, small intestine, mammary gland, and (in s  
F:43-479/Region: EGF precursor long repeat <LRL>  
F:46-85/Domain: LDL receptor YWTD-containing repeat homology <YW01>  
F:86-127/Domain: LDL receptor YWTD-containing repeat homology <YW02>  
F:128-169/Domain: LDL receptor YWTD-containing repeat homology <YW03>  
F:170-211/Domain: LDL receptor YWTD-containing repeat homology <YW04>  
F:212-256/Domain: LDL receptor YWTD-containing repeat homology <YW05>  
F:257-301/Domain: LDL receptor YWTD-containing repeat homology <YW06>  
F:318-354/Domain: EGF homology <EG1>



F:360-395/Domain: EGF homology <EG2>  
 F:401-436/Domain: EGF homology <EG3>  
 F:439-476/Domain: EGF homology <EG4>  
 F:480-954/Region: EGF precursor long repeat <LR2>  
 F:483-523/Domain: LDL receptor WYTD-containing repeat homology <YW07>  
 F:524-566/Domain: LDL receptor WYTD-containing repeat homology <YW08>  
 F:567-609/Domain: LDL receptor WYTD-containing repeat homology <YW09>  
 F:610-653/Domain: LDL receptor WYTD-containing repeat homology <YW10>  
 F:654-694/Domain: LDL receptor WYTD-containing repeat homology <YW11>  
 F:695-737/Domain: LDL receptor WYTD-containing repeat homology <YW12>  
 F:745-780/Domain: EGF homology <EG5>  
 F:835-868/Domain: EGF homology <EG6>  
 F:874-910/Domain: EGF homology <EG7>  
 F:916-951/Domain: EGF homology <EG8>  
 F:971-1023/Product: epidermal growth factor #status experimental <EGF>  
 F:976-1012/Domain: EGF homology <EG9>  
 F:1033-1057/Domain: transmembrane #status predicted <TM>  
 F:1058-1207/Domain: intracellular #status predicted <INT>  
 F:318-330,325-339,341-354,360-371,367-380,382-395,401-412,408-421,423-436,439-451,447-461  
 F:976-990,984-1001,1003-1012/Disulfide bonds: #status experimental

Query Match 81.7%; Score 49; DB 1; Length 1207;  
 Best Local Similarity 70.0%; Pred. No. 1.5;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10  
 ||:|:| |:|:|  
 Db 1003 CVWGIIGERC 1012

RESULT 11  
 A60979  
 versican precursor - human  
 N:Alternate names: chondroitin sulfate proteoglycan 2; chondroitin sulfate proteoglycan  
 N:Contents: glial hyaluronate-binding protein  
 C:Species: Homo sapiens (man)  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 19-Jan-2001  
 C:Accession: S06014; S43921; A60979; A30358; A29348; A45131; I54179  
 R:Zimmermann, D.R.; Ruoslahti, E.  
 EMBO J. 8, 2975-2981, 1989  
 A:Title: Multiple domains of the large fibroblast proteoglycan, versican.  
 A:Reference number: S06014; MUID:90059882  
 A:Accession: S06014  
 A:Molecule type: mRNA  
 A:Residues: 1-2409 <ZIM>  
 A:Cross-references: GB:X15998; NID:g37662; PIDN:CAA34128.1; PID:g37663  
 R:Yao, L.Y.; Moody, C.; Schoenherr, E.; Wight, T.N.; Sandell, L.J.  
 Matrix Biol. 14, 213-225, 1994  
 A:Title: Identification of the proteoglycan versican in aorta and smooth muscle cells by  
 A:Reference number: S43921; MUID:95005762  
 A:Accession: S43921  
 A:Molecule type: mRNA  
 A:Residues: 208-440;1094-1385;1910-2246 <YAO>  
 R:Bignami, A.; Lane, W.S.; Andrews, D.; Dahl, D.  
 Brain Res. Bull. 22, 67-70, 1989  
 A:Title: Structural similarity of hyaluronate binding proteins in brain and cartilage.  
 A:Reference number: A60979; MUID:89229983  
 A:Accession: A60979  
 A:Molecule type: protein  
 A:Residues: 171-210;289-303 <BIG>  
 R:Perides, G.; Lane, W.S.; Andrews, D.; Dahl, D.; Bignami, A.  
 J. Biol. Chem. 264, 5981-5987, 1989  
 A:Title: Isolation and partial characterization of a glial hyaluronate-binding protein.  
 A:Reference number: A30358; MUID:89174663  
 A:Accession: A30358  
 A:Molecule type: protein  
 A:Residues: 24-50;80-87,'D',89-119;128-155;167-218;229-259,'IR';261-268;277-283,'G',285-  
 R:Krusius, T.; Gehlsen, K.R.; Ruoslahti, E.  
 J. Biol. Chem. 262, 13120-13125, 1987  
 A:Title: A fibroblast chondroitin sulfate proteoglycan core protein contains lectin-like  
 A:Reference number: A29348; MUID:88007514

A:Accession: A29348  
 A:Molecule type: mRNA  
 A:Residues: 1725,'V',1727-2409 <KRU>  
 A:Cross-references: GB:J02814  
 R:Perides, G.; Rahemtulla, F.; Lane, W.S.; Asher, R.A.; Bignami, A.  
 J. Biol. Chem. 267, 23883-23887, 1992  
 A:Title: Isolation of a large aggregating proteoglycan from human brain.  
 A:Reference number: A45131; MUID:93054750  
 A:Contents: brain  
 A:Accession: A45131  
 A:Molecule type: protein  
 A:Residues: 21-22,'X',24-37 <PE2>  
 A:Experimental source: brain  
 A:Note: Sequence extracted from NCBI backbone (NCBIP:118884)  
 R:Tozzo, R.V.; Naso, M.F.; Cannizzaro, L.A.; Wasmuth, J.J.; McPherson, J.D.  
 Genomics 14, 845-851, 1992  
 A:Title: Mapping of the versican proteoglycan gene (CSPG2) to the long arm of human c  
 A:Reference number: I54179; MUID:93122792  
 A:Accession: I54179  
 A:Status: translated from GB/EMBL/DBDJ  
 A:Molecule type: DNA  
 A:Residues: 251-347 <RES>  
 A:Cross-references: GB:S52488; NID:g263313; PIDN:AAB24878.1; PID:g263314  
 C:Genetics:  
 A:Gene: GDB:CSPG2  
 A:Cross-references: GDB:I127873; OMIM:118661  
 A:Map position: 5q12-5q14  
 C:Superfamily: versican; C-type lectin homology; complement factor H repeat homology;  
 F:1-20/Domain: signal sequence #status predicted <SIG>  
 F:21-2409/Product: proteoglycan 24K core protein #status predicted <MAT>  
 F:167-244/Domain: link protein repeat homology <LNK1>  
 F:265-346/Domain: link protein repeat homology <LNK2>  
 F:559-1654/Domain: chondroitin sulfate attachment #status predicted <GAG>  
 F:2106-2137/Domain: EGF homology <EG1>  
 F:2144-2175/Domain: EGF homology <EG2>  
 F:2182-2302/Domain: C-type lectin homology <LCH>  
 F:2309-2365/Domain: complement factor H repeat homology <FHD>

Query Match 81.7%; Score 49; DB 1; Length 2409;  
 Best Local Similarity 80.0%; Pred. No. 2.7;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10  
 ||:|:|:|:|  
 Db 2128 CVPGYSGDQC 2137

RESULT 12  
 T30201  
 Notch homolog protein - sea squirt (Halocynthia roretzi)  
 C:Species: Halocynthia roretzi  
 C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000  
 C:Accession: T30201  
 R:Horii, S.; Saitoh, T.; Matsumoto, M.; Makabe, K.W.; Nishida, H.  
 Dev. Genes Evol. 207, 371-380, 1997  
 A:Title: Notch homologue from Halocynthia roretzi is preferentially expressed in the  
 A:Reference number: Z20775  
 A:Accession: T30201  
 A:Status: preliminary; translated from GB/EMBL/DBDJ  
 A:Molecule type: mRNA  
 A:Residues: 1-2352 <HOR>  
 A:Cross-references: EMBL:AB001327; NID:dl204472; PID:dl026501; PIDN:BAA25571.1  
 C:Genetics:  
 A:Gene: Notch

Query Match 80.0%; Score 48; DB 2; Length 2352;  
 Best Local Similarity 80.0%; Pred. No. 4;  
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10  
 ||:|:|:|:|

Db 367 CVAGYSGPRC 376

## RESULT 13

T09070  
 C:Species: Mus musculus (house mouse)  
 C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jan-2000  
 C:Accession: T09070  
 R:Rowen, L.; Mahairas, G.; Qin, S.; Ahearn, M.E.; Dankers, C.; Lasky, S.; Loretz, C.; submitted to the EMBL Data Library, October 1997  
 A:Description: Sequence of the mouse major histocompatibility locus class III region.  
 A:Reference number: Z16543  
 A:Accession: T09070  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-4006 <ROW>  
 A:Cross-references: EMBL:AF030001; NID:g2564945; PID:g2564958  
 C:Genetics: TNX  
 A:Gene: TNX  
 A:Map position: 17  
 A:Introns: 124/1; 735/1; 773/3; 826/1; 914/1; 1037/1; 1135/1; 1232/1; 1329/1; 1440/1; 1501/1; 3115/1; 3208/1; 3302/1; 3405/1; 3517/1; 3558/1; 3606/1; 3646/1; 3694/1; 3737/3; 3742/2-448/Domain: EGF homology <EGF>  
 F:422-448/Domain: EGF homology <EGF>  
 F:826-906/Domain: fibronectin type III repeat homology <3FR>  
 F:3789-3997/Domain: fibronectin type III repeat homology <3FR>  
 F:3789-3997/Domain: fibronectin type III repeat homology <3FR>

## Query Match

Best Local Similarity 80.0%; Score 48; DB 2; Length 4006;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10  
 Db 532 CAVGYSGDRC 541

## RESULT 14

S17294  
 C:Species: Sus scrofa domestica (domestic pig)  
 C:Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 16-Jul-1999  
 C:Accession: S17294  
 R:Pascall, J.C.; Jones, D.S.C.; Doel, S.M.; Clements, J.M.; Hunter, M.; Fallon, T.; Edwa  
 J. Mol. Endocrinol. 6, 63-70, 1991  
 A:Title: Cloning and characterization of a gene encoding pig epidermal growth factor.  
 A:Reference number: S17294; MUID:91197366  
 A:Accession: S17294  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-53 <PAS>  
 A:Cross-references: EMBL:X59516; NID:g1940; PIDN:CAA42102.1; PID:g938287  
 C:Superfamily: epidermal growth factor precursor; EGF homology; LDL receptor YWTD-conta  
 F:6-42/Domain: EGF homology <EG9>

## Query Match

Best Local Similarity 76.7%; Score 46; DB 2; Length 53;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10  
 Db 33 CVFYVGERC 42

## RESULT 15

B69300  
 C:Species: Archaeoglobus fulgidus  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
 C:Accession: B69300

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dcd  
 ; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E  
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
 Nature 390, 364-370, 1997  
 A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artlach, P.; Kaine, B.P.; Sykes,  
 Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arc  
 A:Reference number: A69250; MUID:98049343  
 A:Accession: B69300  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-57 <KLE>  
 A:Cross-references: GB:AE001076; GB:AE000782; NID:g2689399; PIDN:AAB90836.1; PID:g268

Query Match  
 Best Local Similarity 76.7%; Score 46; DB 2; Length 57;  
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10  
 Db 23 CIMGYTGKRC 32

Search completed: August 15, 2001, 10:54:15  
 Job time: 168 sec

us-09-673-785a-2.rpt

Wed Aug 15 10:57:45 2001

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2001, 10:51:28 ; Search time 12.86 Seconds  
(without alignments)  
26.637 Million cell updates/sec

Title: US-09-673-785a-2  
Perfect score: 60  
Sequence: 1 CVIGSGDRC 10

Scoring table:  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	60	100.0	1217	1	EGF_MOUSE
2	55	91.7	2139	1	CRB_DROME
3	50	83.3	756	1	NRG2_MOUSE
4	50	83.3	1133	1	EGF_RAT
5	49	81.7	862	1	PGCV_MACNE
6	49	81.7	1207	1	EGF_HUMAN
7	49	81.7	3396	1	PGCV_HUMAN
8	46	76.7	53	1	EGF_PIG
9	45	75.0	3358	1	PGCV_MOUSE
10	44	73.3	3106	1	LMA2_MOUSE
11	44	73.3	3672	1	LML2_CAEEL
12	43	71.7	1064	1	FBP1_STRPU
13	43	71.7	3097	1	CADN_DROME
14	43	71.7	3110	1	LMA2_HUMAN
15	42	70.0	833	1	DL_DROME
16	42	70.0	883	1	PGCB_RAT
17	42	70.0	2703	1	NOTC_DROME
18	41.5	69.2	412	1	YNPL_CAEEL
19	41	68.3	169	1	EREG_HUMAN
20	41	68.3	183	1	YRF3_SHIEL
21	41	68.3	230	1	SPIT_DROME
22	41	68.3	1790	1	LMB1_DROME
23	41	68.3	1964	1	NTC4_MOUSE
24	41	68.3	2476	1	ZAN_PIG
25	41	68.3	4289	1	TENX_HUMAN
26	40	66.7	473	1	FP2_MYTGA
27	40	66.7	768	1	ITB8_RABIT
28	40	66.7	769	1	ITB8_HUMAN
29	40	66.7	2318	1	NTC3_MOUSE
30	40	66.7	2444	1	NTC1_HUMAN
31	40	66.7	2531	1	NTC1_RAT
32	40	66.7	3084	1	LMA1_MOUSE
33	40	66.7	3635	1	LMA5_MOUSE

34 40 66.7 4544 1 LRPI\_HUMAN  
35 39 65.0 208 1 HBGF\_RAT  
36 39 65.0 427 1 MFGM\_RAT  
37 39 65.0 463 1 MFGM\_MOUSE  
38 39 65.0 477 1 URTL\_DESRO  
39 39 65.0 723 1 DILLI\_HUMAN  
40 39 65.0 883 1 PGCB\_MOUSE  
41 39 65.0 912 1 PGCB\_BOVIN  
42 39 65.0 1408 1 SERR\_DROME  
43 39 65.0 2437 1 NOTC\_BRARE  
44 39 65.0 2531 1 NTC1\_MOUSE  
45 39 65.0 3075 1 LMA1\_HUMAN

## ALIGNMENTS

RESULT 1  
EGF\_MOUSE  
AC P01132;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE PRO-EPIDERMAL GROWTH FACTOR PRECURSOR (EGF) [CONTAINS: EPIDERMAL GROWTH FACTOR]  
DE GROWTH FACTOR]  
GN EGF.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=83223630; PubMed=6602382;  
RA Scott J., Urdea M., Quiroga M., Sanchez-Pescador R., Fong N.M., Selby M., Rutter W.J., Bell G.I.;  
RT "Structure of a mouse submaxillary messenger RNA encoding epidermal growth factor and seven related proteins.";  
RL Science 221:236-240(1983).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=83219309; PubMed=6304537;  
RA Gray A., Dull T.J., Ullrich A.;  
RT "Nucleotide sequence of epidermal growth factor cDNA predicts a 128,000-molecular weight protein precursor.";  
RL Nature 303:722-725(1983).  
RN [3]  
RP SEQUENCE OF 977-1029.  
RX MEDLINE=73048516; PubMed=4636327;  
RA Savage C.R. Jr., Inagami T., Cohen S.;  
RT "The primary structure of epidermal growth factor.";  
RL J. Biol. Chem. 247:7612-7621(1972).  
RN [4]  
RP DISULFIDE BONDS.  
RX MEDLINE=74025498; PubMed=4750422;  
RA Savage C.R. Jr., Hash J.H., Cohen S.;  
RT "Epidermal growth factor. Location of disulfide bonds.";  
RL J. Biol. Chem. 248:7669-7672(1973).  
RN [5]  
RP STRUCTURE BY NMR OF 977-1029.  
RX MEDLINE=92118798; PubMed=1731873;  
RA Montellone G.T., Wuehrich K., Burgess A.W., Nice E.C., Wagner G., Gibson K.D., Scheraga H.A.;  
RT "Solution structure of murine epidermal growth factor determined by NMR spectroscopy and refined by energy minimization with restraints.";  
RL Biochemistry 31:236-249(1992).  
RN [6]  
RP STRUCTURE BY NMR OF 977-1029.  
RX MEDLINE=93075811; PubMed=1445923;  
RA Kohda D., Inagaki F.;  
RT "Three-dimensional nuclear magnetic resonance structures of mouse epidermal growth factor in acidic and physiological pH solutions.";

Q07954 homo sapien  
Q06175 rattus norv  
P70490 rattus norv  
P21956 mus musculu  
P98119 desmodus ro  
Q00548 homo sapien  
Q61361 mus musculu  
Q28062 bos taurus  
P18168 drosophila  
P46530 brachydanio  
Q01705 mus musculu  
P25391 homo sapien

CRB.  
Drosophila melanogaster (Fruit fly).  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
NCBI\_TaxID=7227;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=OREGON-R; TISSUE=Embryo;  
MEDLINE=90263104; PubMed=2344615;  
Tepass U., Theres C., Knust E.;  
"Crumbs encodes an EGF-like protein expressed on apical membranes of  
Drosophila epithelial cells and required for organization of  
epithelia.";  
Cell 61:787-799(1990).  
[2]  
SEQUENCE OF 1663-1955 FROM N.A.  
TISSUE=Embryo;  
MEDLINE=87218537; PubMed=3107986;  
Knust E., Dietrich U., Tepass U., Bremer K.A., Weigel D.,  
Vaessen H., Campos-Ortega J.A.;



FT DISULFID 1905 1914 BY SIMILARITY.  
 FT DISULFID 1919 1930 BY SIMILARITY.  
 FT DISULFID 1924 1939 BY SIMILARITY.  
 FT DISULFID 1941 1950 BY SIMILARITY.  
 FT DISULFID 1957 1968 BY SIMILARITY.  
 FT DISULFID 1962 1977 BY SIMILARITY.  
 FT DISULFID 1979 1988 BY SIMILARITY.  
 FT DISULFID 1995 2008 BY SIMILARITY.  
 FT DISULFID 2002 2017 BY SIMILARITY.  
 FT DISULFID 2019 2028 BY SIMILARITY.  
 FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 400 400 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 565 565 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 736 736 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 746 746 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 860 860 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 884 884 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 976 976 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1102 1102 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1138 1138 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 Query Match 91.7%; Score 55; DB 1; Length 2139;  
 Best Local Similarity 80.0%; Pred. No. 0.038;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 CVIGYSGDRC 10  
 Db 571 CAVGYSGDRC 580  
 RESULT 3  
 ID NRG2\_MOUSE STANDARD; PRT: 756 AA.  
 AC P56974;  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DE PRO-NEUREGULIN-2 PRECURSOR (PRO-NRG2) [CONTAINS: NEUREGULIN-2 (NRG-2)  
 DE (DIVERGENT OF NEUREGULIN 1) (DON-1)].  
 GN NRG2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS NRG2-5; NRG2-10 AND NRG2-16A).  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RX MEDLINE=97311396; PubMed=9168115;  
 RA Caraway K.L. III, Weber J.L., Unger M.J., Ledesma J., Yu N.,  
 RA Gassmann M., Lai C.,  
 RT "Neuregulin-2, a new ligand of ErbB3/ErbB4-receptor tyrosine  
 RT kinases".  
 RL Nature 387:512-516(1997).  
 RN [2]  
 RP SEQUENCE OF 150-756 FROM N.A. (ISOFORMS DON-1M AND DON-1S).  
 RC TISSUE=Choroid plexus;  
 RX MEDLINE=97342638; PubMed=9199335;  
 RA Busfield S.J., Michnick D.A., Chickering T.W., Revett T.L., Ma J.,  
 RA Woolf E.A., Conrath C.A., Dussault B.J., Woolf J., Goodearl A.D.J.,  
 RA Gearing D.P.;  
 RT "Characterization of a neuregulin-related gene, Don-1, that is highly  
 RT expressed in restricted regions of the cerebellum and hippocampus".  
 RL Mol. Cell. Biol. 17:4007-4014(1997).  
 CC -!- FUNCTION: DIRECT LIGAND FOR ERBB3 AND ERBB4 TYROSINE KINASE  
 CC RECEPTORS. CONCOMITANTLY RECRUITS ERBB1 AND ERBB2 CORECEPTORS,  
 CC RESULTING IN LIGAND-STIMULATED TYROSINE PHOSPHORYLATION AND

CC ACTIVATION OF THE ERBB RECEPTORS. MAY ALSO PROMOTE THE  
 CC HETERODIMERIZATION WITH THE EGF RECEPTOR.  
 CC -!- SUBCELLULAR LOCATION: EXISTS AS AN TYPE I MEMBRANE PROTEIN AND AS  
 CC A PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE  
 CC MEMBRANE-BOUND FORM DOES NOT SEEM TO BE ACTIVE (BY SIMILARITY).  
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST 4 ISOFORMS; DON-1M, DON-1S/NRG2-5,  
 CC NRG2-10 AND NRG2-16A (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE  
 CC SPLICING.  
 CC -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN THE BRAIN, WITH LOWER  
 CC LEVELS IN THE LUNG. IN THE CEREBELLUM, FOUND IN GRANULE AND  
 CC PURKINJE CELLS.  
 CC -!- DOMAIN: THE CYTOPLASMIC DOMAIN MAY BE INVOLVED IN THE REGULATION  
 CC OF TRAFFICKING AND PROTEOLYTIC PROCESSING. REGULATION OF THE  
 CC PROTEOLYTIC PROCESSING INVOLVES INITIAL INTRACELLULAR DOMAIN  
 CC DIMERIZATION (BY SIMILARITY).  
 CC -!- DOMAIN: ERBB RECEPTOR BINDING IS ELICITED ENTIRELY BY THE EGF-LIKE  
 CC DOMAIN (BY SIMILARITY).  
 CC -!- PTM: PROTEOLYTIC CLEAVAGE CLOSE TO THE PLASMA MEMBRANE ON THE  
 CC EXTERNAL FACE LEADS TO THE RELEASE OF THE SOLUBLE GROWTH FACTOR  
 CC FORM (BY SIMILARITY).  
 CC -!- PTM: EXTENSIVE GLYCOSYLATION PRECEDES THE PROTEOLYTIC CLEAVAGE (BY  
 CC SIMILARITY).  
 CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.  
 CC -!- SIMILARITY: BELONGS TO THE NEUREGULIN FAMILY.  
 CC MGD; MGI:1096828; Nrg2.  
 DR InterPro; IPR000561; -;  
 DR InterPro; IPR003006; -;  
 DR Pfam; PF000047; EGF; 1.  
 DR Pfam; PF00008; EGF; 1.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 KW Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;  
 KW Transmembrane; Multigene family; Alternative splicing.  
 FT PROPEP 1 19  
 FT CHAIN 20 756  
 FT DOMAIN 20 314  
 FT TRANSMEM 20 315  
 FT DOMAIN 316 336  
 FT DOMAIN 337 756  
 FT DOMAIN 158 226  
 FT DOMAIN 238 248  
 FT DOMAIN 249 290  
 FT DOMAIN 627 633  
 FT DISULFID 165 219  
 FT DISULFID 253 267  
 FT DISULFID 261 278  
 FT DISULFID 280 289  
 FT CARBOHYD 55 55  
 FT CARBOHYD 186 186  
 FT CARBOHYD 254 254  
 FT CARBOHYD 296 296  
 FT VARSPLIC 280 280  
 FT VARSPLIC 281 281  
 FT VARSPLIC 282 330  
 FT VARSPLIC 331 756  
 FT VARSPLIC 282 307  
 SQ SEQUENCE 756 AA; 82213 MW; 51D85DC918BE678E CRC64;  
 Query Match 83.3%; Score 50; DB 1; Length 756;  
 Best Local Similarity 70.0%; Pred. No. 0.11;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 CVIGYSGDRC 10  
 Db 280 CPVGYTGDCRC 289  
 RESULT 4





OS Macaca nemestrina (Pig-tailed macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;  
 OC Cercopitheciinae; Macaca.  
 OX NCBI\_TaxID=9545;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Aortic smooth muscle;  
 RX MEDLINE=95005762; PubMed=7921538;  
 RA Yao L.Y., Moody C., Schoenherr E., Wight T.N., Sandell L.J.;  
 RT "Identification of the proteoglycan versican in aorta and smooth  
 muscle cells by DNA sequence analysis, in situ hybridization and  
 immunohistochemistry.";  
 RL Matrix Biol. 14:213-225(1994).  
 CC -!- FUNCTION: MAY PLAY A ROLE IN INTERCELLULAR SIGNALING AND IN  
 CC CONNECTING CELLS WITH THE EXTRACELLULAR MATRIX. MAY TAKE PART IN  
 CC THE REGULATION OF CELL MOTILITY, GROWTH AND DIFFERENTIATION. BINDS  
 CC HYALURONAN.  
 CC -!- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY  
 CC SIMILARITY).  
 CC -!- DEVELOPMENTAL STAGE: DISAPPEARS AFTER THE CARTILAGE DEVELOPMENT  
 CC (BY SIMILARITY).  
 CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.  
 CC -!- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: S72412; AAA65593.2; -  
 DR EMBL: S72413; AAA65594.2; -  
 DR EMBL: S72414; AAA65595.2; -  
 DR HSP: P01132; 1EPH.  
 DR InterPro: IPR000152; -  
 DR InterPro: IPR000538; -  
 DR InterPro: IPR000561; -  
 DR InterPro: IPR001304; -  
 DR InterPro: IPR001881; -  
 DR Pfam: PF00008; EGF; 2.  
 DR Pfam: PF00193; Xlink; 2.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 1.  
 DR PROSITE: PS00022; EGF\_1; 2.  
 DR PROSITE: PS01186; EGF\_2; 1.  
 DR PROSITE: PS01187; EGF\_CA; 1.  
 DR PROSITE: PS01241; LINK; 1.  
 DR PROSITE: PS00615; C-TYPE\_LECTIN\_1; PARTIAL.  
 DR PROSITE: PS50041; C-TYPE\_LECTIN\_2; 1.  
 KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Repeat;  
 KW EGF-like domain; Calcium.  
 FT NON\_TER 1  
 FT DOMAIN <1 37 LINK 1.  
 FT DOMAIN 58 139 LINK 2.  
 FT DOMAIN 141 >233 GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT  
 FT DOMAIN) DOMAIN).  
 FT NON\_CONS 233 234 GLUCOSAMINOGLYCAN ATTACHMENT DOMAIN,  
 FT DOMAIN <234 >525 SIMILAR TO CHONDROITIN SULFATE ATTACHMENT  
 FT DOMAIN SITE IN COLLAGEN TYPE IX (BY SIMILARITY).  
 FT  
 FT NON\_CONS 525 526 EGF-LIKE 1.  
 FT DOMAIN 718 754 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 756 792 C-TYPE LECTIN.  
 FT DOMAIN 795 >862 BY SIMILARITY.  
 FT DISULFID 63 138 BY SIMILARITY.  
 FT DISULFID 87 108 BY SIMILARITY.  
 FT DISULFID 722 733 BY SIMILARITY.  
 FT DISULFID 727 742 BY SIMILARITY.  
 FT DISULFID 744 753 BY SIMILARITY.

FT DISULFID 760 771 BY SIMILARITY.  
 FT DISULFID 765 780 BY SIMILARITY.  
 FT DISULFID 782 791 BY SIMILARITY.  
 FT CARBOHYD 123 123 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 204 204 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 425 425 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 513 513 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 563 563 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 696 696 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT NON\_TER 862 862 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 862 AA; 95593 MW; ASD5F6153A74BB39 CRC64;  
 Query Match 81.7%; Score 49; DB 1; Length 862;  
 Best Local Similarity 80.0%; Pred. No. 0.2;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CVIGYSGDRC 10  
 Db 744 CVPYSGDQC 753  
 II IIIII:I  
 RESULT 6  
 EGF\_HUMAN  
 ID EGF\_HUMAN STANDARD; PRT; 1207 AA.  
 AC P01133;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DE 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE PRO-EPIDERMAL GROWTH FACTOR PRECURSOR (EGF) [CONTAINS: EPIDERMAL  
 DE GROWTH FACTOR (UROGASTRONE)].  
 GN EGF.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=87066721; PubMed=3491360;  
 RA Bell G.I., Fong N.M., Stempien M.M., Wormsted M.A., Caput D.,  
 RA Ku L., Urdea M.S., Rall L.B., Sanchez-Pescador R.;  
 RT "Human epidermal growth factor precursor: cDNA sequence, expression  
 in vitro and gene organization.";  
 RL Nucleic Acids Res. 14:8427-8446(1986).  
 RN [2]  
 RP SEQUENCE OF 971-1023.  
 RX MEDLINE=77117897; PubMed=300079;  
 RA Gregory H., Preston B.M.;  
 RT "The primary structure of human urogastrone.";  
 RL Int. J. Pept. Protein Res. 9:107-118(1977).  
 RN [3]  
 RP SEQUENCE OF 971-1023.  
 RX MEDLINE=89391964; PubMed=2789514;  
 RA Furuya M., Akashi S., Hirayama K.;  
 RT "The primary structure of human EGF produced by genetic engineering,  
 studied by high-performance tandem mass spectrometry.";  
 RL Biochem. Biophys. Res. Commun. 163:1100-1106(1989).  
 RN [4]  
 RP STRUCTURE BY NMR OF EGF.  
 RX MEDLINE=92395667; PubMed=1522591;  
 RA Hommel U., Harvey T.S., Driscoll P.C., Campbell I.D.;  
 RT "Human epidermal growth factor. High resolution solution structure  
 and comparison with human transforming growth factor alpha.";  
 RL J. Mol. Biol. 227:271-282(1992).  
 CC -!- FUNCTION: THE GROWTH FACTOR STIMULATES THE GROWTH OF VARIOUS  
 CC EPIDERMAL AND EPITHELIAL TISSUES IN VIVO AND IN VITRO AND OF SOME  
 CC FIBROBLASTS IN CELL CULTURE.  
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -!- SIMILARITY: CONTAINS 9 EGF-LIKE DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

EMBL; X04571; CAA28240.1; --  
 PIR; A25531; EGHU.  
 HSP; P01132; 1EPH.  
 MIM; 131530; --  
 InterPro; IPR000033; --  
 InterPro; IPR000152; --  
 InterPro; IPR000561; --  
 InterPro; IPR001336; --  
 InterPro; IPR001881; --  
 Pfam; PF00008; EGF; 9.  
 Pfam; PF00058; Idl\_recept\_b; 7.  
 PRINTS; PR00009; EGTGF.  
 PROSITE; PS00010; ASX\_HYDROXYL; 3.  
 PROSITE; PS00022; EGF\_1; 1.  
 PROSITE; PS01186; EGF\_2; 7.  
 PROSITE; PS01187; EGF\_CA; 3.  
 EGF-like domain; Repeat; Growth factor; Transmembrane; Glycoprotein;  
 Signal; Polymorphism.  
 SIGNAL 1 22  
 CHAIN 23 1207  
 CHAIN 971 1023  
 DOMAIN 23 1032  
 TRANSMEM 1033 1053  
 DOMAIN 1054 1207  
 DOMAIN 314 355  
 DOMAIN 356 396  
 DOMAIN 397 437  
 DOMAIN 435 477  
 DOMAIN 741 781  
 DOMAIN 831 869  
 DOMAIN 870 911  
 DOMAIN 912 952  
 DOMAIN 972 1013  
 DISULFID 318 330  
 DISULFID 325 339  
 DISULFID 341 354  
 DISULFID 360 371  
 DISULFID 367 380  
 DISULFID 382 395  
 DISULFID 401 412  
 DISULFID 408 421  
 DISULFID 423 436  
 DISULFID 439 451  
 DISULFID 447 461  
 DISULFID 463 476  
 DISULFID 745 756  
 DISULFID 752 765  
 DISULFID 767 780  
 DISULFID 835 846  
 DISULFID 840 855  
 DISULFID 857 868  
 DISULFID 874 888  
 DISULFID 881 897  
 DISULFID 899 910  
 DISULFID 916 929  
 DISULFID 923 938  
 DISULFID 940 951  
 DISULFID 976 990  
 DISULFID 984 1001  
 DISULFID 1003 1012  
 CARBOHYD 38 304  
 CARBOHYD 104 104  
 CARBOHYD 117 117  
 CARBOHYD 148 148  
 CARBOHYD 324 324

FT CARBOHYD 404 404 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 596 596 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 815 815 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 926 926 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARIANT 708 708 I -> M.  
 FT SEQUENCE 1207 AA; 133945 MW; D627DC828EF782E9 CRC64;  
 SQ  
 Query Match 81.7%; Score 49; DB 1; Length 1207;  
 Best Local Similarity 70.0%; Pred.No. 0.28;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CVIGSYGDR 10  
 II:II I:II  
 Db 1003 CVVGYIGERC 1012  
 RESULT 7  
 PGCV\_HUMAN  
 ID PGCV\_HUMAN STANDARD; PRT; 3396 AA.  
 AC P13611; P20754; Q13010; Q13189; Q15123;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE VERSICAN CORE PROTEIN PRECURSOR (LARGE FIBROBLAST PROTEOGLYCAN)  
 DE (CHONDROITIN SULFATE PROTEOGLYCAN CORE PROTEIN 2) (GLIAL HYALURONATE-  
 BINDING PROTEIN) (GHAP).  
 GN CSPG2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (VARIANT V0).  
 RX MEDLINE=95105188; PubMed=7528742;  
 RA Naso M.F., Zimmermann D.R., Iozzo R.V.;  
 RT "Characterization of the complete genomic structure of the human  
 versican gene and functional analysis of its promoter.";  
 RL J. Biol. Chem. 269:32999-33008(1994).  
 RN [2]  
 RP SEQUENCE OF 1-347 AND 1336-3396 FROM N.A. (VARIANT V1).  
 RC TISSUE=Placenta;  
 RX MEDLINE=90059882; PubMed=2583089;  
 RA Zimmermann D.R., Ruoslahti E.;  
 RT "Multiple domains of the large fibroblast proteoglycan, versican.";  
 RL EMBO J. 8:2975-2981(1989).  
 RN [3]  
 RP SEQUENCE OF 1-1335 AND 3090-3396 FROM N.A. (VARIANT V2).  
 RC TISSUE=Glial tumor;  
 RX MEDLINE=95105187; PubMed=7806529;  
 RA Dours-Zimmermann M.T., Zimmermann D.R.;  
 RT "A novel glycosaminoglycan attachment domain identified in two  
 alternative splice variants of human versican.";  
 RL J. Biol. Chem. 269:32992-32998(1994).  
 RN [4]  
 RP SEQUENCE OF 2711-3396 FROM N.A.  
 RC TISSUE=Lung fibroblast;  
 RX MEDLINE=88007514; PubMed=2820964;  
 RA Krusius T., Gehlsen K.R., Ruoslahti E.;  
 RT "A fibroblast chondroitin sulfate proteoglycan core protein contains  
 lectin-like and growth factor-like sequences.";  
 RL J. Biol. Chem. 262:13120-13125(1987).  
 RN [5]  
 RP SEQUENCE OF 251-347 FROM N.A.  
 RX MEDLINE=93122792; PubMed=1478664;  
 RA Iozzo R.V., Naso M.F., Cannizzaro L.A., Wasmuth J.J.,  
 RA McPherson J.D.;  
 RT "Mapping of the versican proteoglycan gene (CSPG2) to the long arm of  
 human chromosome 5 (5q12-5q14).";  
 RL Genomics 14:845-851(1992).  
 RN [6]  
 RP SEQUENCE OF 1-347 AND 3090-3396 FROM N.A.

DR	PRINTS: PR00010; EGFBL00D.
DR	PROSITE; PS00010; ASX_HYDROXYL; 1.
DR	PROSITE; PS00022; EGF_1; 2.
DR	PROSITE; PS01186; EGF_2; 1.
DR	PROSITE; PS01187; EGF_CA; 1.
DR	PROSITE; PS01241; LINK; 2.
DR	PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR	PROSITE; PS00041; C_TYPE_LECTIN_2; 1.
KW	Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW	Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW	Hyaluronic acid; Alternative splicing.
FT	SIGNAL 1 20 POTENTIAL.
FT	CHAIN 21 3396 VERSICAN CORE PROTEIN.
FT	DOMAIN 37 137 IG-LIKE V-TYPE DOMAIN.
FT	DOMAIN 167 244 LINK 1.
FT	DOMAIN 265 346 LINK 2.
FT	DOMAIN 348 1335 GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT DOMAIN).
FT	DOMAIN 1336 3089 GAG-BETA.
FT	DOMAIN 3089 3125 EGF-LIKE 1.
FT	DOMAIN 3127 3163 EGF-LIKE 2.
FT	DOMAIN 3166 3292 C-TYPE LECTIN.
FT	DOMAIN 3295 3353 SUSHI.
FT	DISULFID 44 130 BY SIMILARITY.
FT	DISULFID 172 243 BY SIMILARITY.
FT	DISULFID 196 217 BY SIMILARITY.
FT	DISULFID 270 345 BY SIMILARITY.
FT	DISULFID 294 315 BY SIMILARITY.
FT	DISULFID 3093 3104 BY SIMILARITY.
FT	DISULFID 3098 3113 BY SIMILARITY.
FT	DISULFID 3115 3124 BY SIMILARITY.
FT	DISULFID 3131 3142 BY SIMILARITY.
FT	DISULFID 3136 3151 BY SIMILARITY.
FT	DISULFID 3163 3162 BY SIMILARITY.
FT	DISULFID 3169 3180 BY SIMILARITY.
FT	DISULFID 3197 3289 BY SIMILARITY.
FT	DISULFID 3265 3281 BY SIMILARITY.
FT	DISULFID 3296 3339 BY SIMILARITY.
FT	DISULFID 3325 3352 BY SIMILARITY.
FT	CARBOHYD 57 57 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 330 330 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 615 615 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 782 782 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 809 809 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 1332 1332 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 1398 1398 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 1442 1442 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 1468 1468 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 1663 1663 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 1898 1898 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 2179 2179 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 2272 2272 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 2280 2280 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 2360 2360 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 2385 2385 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 2392 2392 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 2436 2436 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 2628 2628 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 2934 2934 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 3067 3067 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 3369 3369 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 3379 3379 N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VARSPPLIC 348 1335 MISSING (IN ISOFORM V1).
FT	VARSPPLIC 1336 3089 MISSING (IN ISOFORM V2).
FT	VARSPPLIC 348 3089 MISSING (IN ISOFORM V3).
FT	CONFLICT 88 88 N -> D (IN REF. 7).
FT	CONFLICT 260 260 K -> I (IN REF. 7).
FT	CONFLICT 274 274 D -> A (IN REF. 6).
FT	CONFLICT 284 284 Q -> G (IN REF. 7).
FT	CONFLICT 348 348 P -> R (IN REF. 6).
FT	CONFLICT 2709 2713 IKAEEA -> EFREY (IN REF. 4).
FT	SEQUENCE 3396 AA; 372815 MW; D174A1BB38304FEC CRC64:

Query Match 81.7%; Score 49; DB 1; Length 3396;  
Best Local Similarity 80.0%; Pred. No. 0.76;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10  
|| |||||:|  
Db 3115 CVPGYSGDQC 3124

## RESULT 8

EGF\_PIG STANDARD; PRT; 53 AA.  
AC Q00968;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE EPIDERMAL GROWTH FACTOR (EGF) (FRAGMENT).  
GN EGF.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
OX NCBI\_TaxID=9823;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Kidney;  
RX MEDLINE=91197366; PubMed=2015058;  
RA Pascall J.C., Jones D.S.C., Doel S.M., Clements J.M., Hunter M.,  
RA Fallon T., Edwards M., Brown K.D.;  
RT "Cloning and characterization of a gene encoding pig epidermal growth factor."  
RL J. Mol. Endocrinol. 6:63-70(1991).  
CC -!- FUNCTION: THE GROWTH FACTOR STIMULATES THE GROWTH OF VARIOUS EPIDERMAL AND EPITHELIAL TISSUES IN VIVO AND IN VITRO AND OF SOME FIBROBLASTS IN CELL CULTURE.  
CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X59516; CAA42102.1; -  
CC HSP; P01132; LEPH.  
CC InterPro: IPR000561; -  
CC Pfam; PF00008; EGF; 1.  
CC PROSITE; PS00022; EGF\_1; 1.  
CC PROSITE; PS01186; EGF\_2; 1.  
KW EGF-like domain; Growth factor.  
FT NON\_TER 1 1  
FT PEPTIDE 1 53 EPIDERMAL GROWTH FACTOR.  
FT DOMAIN 2 43 EGF-LIKE.  
FT DISULFID 6 20 BY SIMILARITY.  
FT DISULFID 14 31 BY SIMILARITY.  
FT DISULFID 33 42 BY SIMILARITY.  
FT NON\_TER 53 53  
SQ SEQUENCE 53 AA; 6149 MW; 74F615B4A05774D4 CRC64;

Query Match 76.7%; Score 46; DB 1; Length 53;  
Best Local Similarity 70.0%; Pred. No. 0.046;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10  
|| |||:|  
Db 33 CVPGYSGDRC 42

## RESULT 9

PGCV\_MOUSE

PGCV\_MOUSE STANDARD; PRT; 3358 AA.  
Q62059; Q62058;  
AC 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE VERSICAN CORE PROTEIN PRECURSOR (LARGE FIBROBLAST PROTEOGLYCAN) (CHONDROITIN SULFATE PROTEOGLYCAN CORE PROTEIN 2) (PG-M).  
GN CSPG2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (VARIANTS V0; V1 AND V2).  
RC STRAIN=C57BL/6, AND SWISS WEBSTER; TISSUE=Brain;  
RX MEDLINE=95122551; PubMed=7822336;  
RA Ito K., Shinomura T., Zako M., Ujita M., Kimata K.;  
RT "Multiple forms of mouse PG-M, a large chondroitin sulfate proteoglycan generated by alternative splicing."  
RL J. Biol. Chem. 270:958-965(1995).  
RN [2]  
RP SEQUENCE OF 1-348 AND 3053-3358 FROM N.A. (VARIANT V3).  
RC STRAIN=C57BL/6;  
RX MEDLINE=95181355; PubMed=7876137;  
RA Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;  
RT "Expression of PG-M(V3), an alternatively spliced form of PG-M without a chondroitin sulfate attachment in region in mouse and human tissues."  
RL J. Biol. Chem. 270:3914-3918(1995).  
CC -!- FUNCTION: MAY PLAY A ROLE IN INTERCELLULAR SIGNALING AND IN CONNECTING CELLS WITH THE EXTRACELLULAR MATRIX. MAY TAKE PART IN THE REGULATION OF CELL MOTILITY, GROWTH AND DIFFERENTIATION. BINDS HYALURONIC ACID.  
CC -!- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX.  
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 4 ISOFORMS; V0 (SHOWN HERE), V1, V2 AND V3; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- TISSUE SPECIFICITY: V2 IS FOUND ONLY IN BRAIN.  
CC -!- DEVELOPMENTAL STAGE: DISAPPEARS AFTER THE CARTILAGE DEVELOPMENT.  
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.  
CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.  
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) REPEAT.  
CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D16263; BAA03796.1; -  
CC EMBL; D28599; -; NOT\_ANNOTATED\_CDS.  
CC EMBL; D32040; BAA06802.1; -  
CC HSP; P00740; IIXA.  
CC MGD; MGI:102889; Cspg2.  
CC InterPro: IPR000152; -  
CC InterPro: IPR000436; -  
CC InterPro: IPR000538; -  
CC InterPro: IPR000561; -  
CC InterPro: IPR001304; -  
CC InterPro: IPR001438; -  
CC InterPro: IPR001881; -  
CC InterPro: IPR003006; -  
CC Pfam; PF00008; EGF; 2.  
CC Pfam; PF00193; xlink; 2.  
CC Pfam; PF00047; ig; 1.  
CC Pfam; PF00059; lectin\_c; 1.  
CC Pfam; PF00084; sushi; 1.  
CC PRINTS; PR00010; EGFBL00D.  
CC PROSITE; PS00010; ASX\_HYDROXYL; 1.

DR PROSITE; PS00022; EGF\_1; 2.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS01187; EGF\_CA; 1.  
 DR PROSITE; PS01241; LINK; 2.  
 DR PROSITE; PS00615; C\_TYPE\_LECTIN\_1; 1.  
 DR PROSITE; PS00041; C\_TYPE\_LECTIN\_2; 1.  
 KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;  
 KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;  
 KW Hyaluronic acid; Alternative splicing.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 21 3358 VERSICAN CORE PROTEIN.  
 FT DOMAIN 37 137 IG-LIKE V-TYPE DOMAIN.  
 FT DOMAIN 167 244 LINK 1.  
 FT DOMAIN 265 334 LINK 2.  
 FT DOMAIN 348 1308 GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT DOMAIN).  
 FT DOMAIN 1309 3052 GAG-BETA.  
 FT DOMAIN 3052 3088 EGF-LIKE 1.  
 FT DOMAIN 3090 3126 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 3129 3255 C-TYPE LECTIN.  
 FT DOMAIN 3258 3316 SUSHI.  
 FT DISULFID 44 130 BY SIMILARITY.  
 FT DISULFID 172 243 BY SIMILARITY.  
 FT DISULFID 196 217 BY SIMILARITY.  
 FT DISULFID 270 333 BY SIMILARITY.  
 FT DISULFID 294 315 BY SIMILARITY.  
 FT DISULFID 3056 3067 BY SIMILARITY.  
 FT DISULFID 3061 3076 BY SIMILARITY.  
 FT DISULFID 3078 3087 BY SIMILARITY.  
 FT DISULFID 3094 3105 BY SIMILARITY.  
 FT DISULFID 3099 3114 BY SIMILARITY.  
 FT DISULFID 3116 3125 BY SIMILARITY.  
 FT DISULFID 3132 3143 BY SIMILARITY.  
 FT DISULFID 3160 3252 BY SIMILARITY.  
 FT DISULFID 3228 3244 BY SIMILARITY.  
 FT DISULFID 3259 3302 BY SIMILARITY.  
 FT DISULFID 3288 3315 BY SIMILARITY.  
 FT CARBOHYD 57 57 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 330 330 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 351 351 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 441 441 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 807 807 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 914 914 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 951 951 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1305 1305 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1372 1372 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1679 1679 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2054 2054 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2244 2244 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2362 2362 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2627 2627 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 3030 3030 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 3332 3332 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 3342 3342 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT VARSPLIC 349 1308 MISSING (IN ISOFORM V1).  
 FT VARSPLIC 1309 3052 MISSING (IN ISOFORM V2).  
 FT VARSPLIC 349 3052 MISSING (IN ISOFORM V3).  
 FT CONFLICT 348 348 P -> R (IN REF. 2).  
 SQ SEQUENCE 3358 AA; 366938 MW; 071B80026BC0762D CRC64;

Query Match 75.0%; Score 45; DB 1; Length 3358;  
 Best Local Similarity 70.0%; Pred No. 4.1;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVTIGSGDRC 10  
 Db 3078 CAPGSGDQC 3087  
 RESULT 10  
 LMA2\_MOUSE  
 ID LMA2\_MOUSE STANDARD; PRT: 3106 AA.

AC Q60675; Q05003; Q64061;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE LAMININ ALPHA-2 CHAIN PRECURSOR (LAMININ M CHAIN) (MEROSIN HEAVY CHAIN).  
 DE CHAIN).  
 GN LAMA2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=Embryo, and Heart;  
 RX MEDLINE=953116259; PubMed=7795883;  
 RA Bernier S.M., Utani A., Sugiyama S., Doi T., Polistina C., Yamada Y.;  
 RA "Cloning and expression of laminin alpha 2 chain (M-chain) in the mouse";  
 RL Matrix Biol. 14:447-455(1995).  
 RN [2]  
 RP SEQUENCE OF 2162-2279 FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Thymus;  
 RX MEDLINE=93346725; PubMed=8345183;  
 RA Chang A.C., Wadsworth S., Colligan J.E.;  
 RA "Expression of merosin in the thymus and its interaction with thymocytes";  
 RL J. Immunol. 151:1789-1801(1993).  
 RN [3]  
 RP SEQUENCE OF 64-281 FROM N.A.  
 RX MEDLINE=95179178; PubMed=7874173;  
 RA Xu H., Wu X.R., Wewer U.M., Engvall E.;  
 RA "Murine muscular dystrophy caused by a mutation in the laminin alpha 2 (Lama2) gene";  
 RL Nat. Genet. 8:297-302(1994).  
 CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.  
 CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END. THE ALPHA-2 CHAIN IS A SUBUNIT OF LAMININ-2 (MEROSIN) AND LAMININ-4 (S-MEROSIN).  
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
 CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR COMPONENT).  
 CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.  
 CC -1- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.  
 CC -1- DISEASE: DEFECTS IN LAMA2 ARE A CAUSE OF MURINE MUSCULAR DYSTROPHY (DY2J).  
 CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).  
 CC -1- SIMILARITY: CONTAINS 17 LAMININ EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.  
 CC -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; U12147; AAC52165.1; -;  
 CC EMBL; X69869; CAA49502.1; -;  
 CC EMBL; S75315; AAB33573.1; -;  
 CC HSP; P02468; IKLO.  
 CC MGI; MGI:99912; Lama2.  
 CC InterPro; IPR000034; -;  
 CC InterPro; IPR000561; -;.

```

DR InterPro; IPR001791; -
DR InterPro; IPR001886; -
DR Pfam; PF00052; laminin_B; 2.
DR Pfam; PF00053; laminin_EGF; 15.
DR Pfam; PF00054; laminin_G; 5.
DR Pfam; PF00055; laminin_Nterm; 1.
DR PRINTS; PR00011; EGF_LAMININ.
DR PROSITE; PS00022; EGF_1; 11.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 14.
KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
FT SIGNAL 1 22
FT CHAIN 23 3106
FT DOMAIN 23 282
FT DOMAIN 283 523
FT DOMAIN 283 339
FT DOMAIN 340 409
FT DOMAIN 410 464
FT DOMAIN 465 513
FT DOMAIN 514 523
FT DOMAIN 524 719
FT DOMAIN 720 1171
FT DOMAIN 720 752
FT DOMAIN 753 802
FT DOMAIN 803 860
FT DOMAIN 861 913
FT DOMAIN 914 962
FT DOMAIN 963 1009
FT DOMAIN 1010 1055
FT DOMAIN 1056 1101
FT DOMAIN 1102 1161
FT DOMAIN 1162 1171
FT DOMAIN 1172 1375
FT DOMAIN 1376 1568
FT DOMAIN 1376 1415
FT DOMAIN 1416 1464
FT DOMAIN 1465 1522
FT DOMAIN 1523 1569
FT DOMAIN 1570 2149
FT DOMAIN 2150 3106
FT DOMAIN 2150 2323
FT DOMAIN 2334 2516
FT DOMAIN 2517 2746
FT DOMAIN 2747 2929
FT DOMAIN 2930 3106
FT DOMAIN 1662 1863
FT DOMAIN 1923 2146
FT DISULFID 283 292
FT DISULFID 285 303
FT DISULFID 305 314
FT DISULFID 317 337
FT DISULFID 340 349
FT DISULFID 342 374
FT DISULFID 377 386
FT DISULFID 389 407
FT DISULFID 410 422
FT DISULFID 412 438
FT DISULFID 440 449
FT DISULFID 452 462
FT DISULFID 465 478
FT DISULFID 467 482
FT DISULFID 484 493
FT DISULFID 496 511
FT DISULFID 753 762
FT DISULFID 755 769
FT DISULFID 772 781
FT DISULFID 784 800
FT DISULFID 803 818
InterPro; IPR001791; -
InterPro; IPR001886; -
Pfam; PF00052; laminin_B; 2.
Pfam; PF00053; laminin_EGF; 15.
Pfam; PF00054; laminin_G; 5.
Pfam; PF00055; laminin_Nterm; 1.
PRINTS; PR00011; EGF_LAMININ.
PROSITE; PS00022; EGF_1; 11.
PROSITE; PS01186; EGF_2; 3.
PROSITE; PS01248; LAMININ_TYPE_EGF; 14.
Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
SIGNAL 1 22
CHAIN 23 3106
DOMAIN 23 282
DOMAIN 283 523
DOMAIN 283 339
DOMAIN 340 409
DOMAIN 410 464
DOMAIN 465 513
DOMAIN 514 523
DOMAIN 524 719
DOMAIN 720 1171
DOMAIN 720 752
DOMAIN 753 802
DOMAIN 803 860
DOMAIN 861 913
DOMAIN 914 962
DOMAIN 963 1009
DOMAIN 1010 1055
DOMAIN 1056 1101
DOMAIN 1102 1161
DOMAIN 1162 1171
DOMAIN 1172 1375
DOMAIN 1376 1568
DOMAIN 1376 1415
DOMAIN 1416 1464
DOMAIN 1465 1522
DOMAIN 1523 1569
DOMAIN 1570 2149
DOMAIN 2150 3106
DOMAIN 2150 2323
DOMAIN 2334 2516
DOMAIN 2517 2746
DOMAIN 2747 2929
DOMAIN 2930 3106
DOMAIN 1662 1863
DOMAIN 1923 2146
DISULFID 283 292
DISULFID 285 303
DISULFID 305 314
DISULFID 317 337
DISULFID 340 349
DISULFID 342 374
DISULFID 377 386
DISULFID 389 407
DISULFID 410 422
DISULFID 412 438
DISULFID 440 449
DISULFID 452 462
DISULFID 465 478
DISULFID 467 482
DISULFID 484 493
DISULFID 496 511
DISULFID 753 762
DISULFID 755 769
DISULFID 772 781
DISULFID 784 800
DISULFID 803 818

```

```

FT DISULFID 805 828
FT DISULFID 831 840
FT DISULFID 843 858
FT DISULFID 861 875
FT DISULFID 863 882
FT DISULFID 885 894
FT DISULFID 897 911
FT DISULFID 914 926
FT DISULFID 916 933
FT DISULFID 935 944
FT DISULFID 947 960
FT DISULFID 963 975
FT DISULFID 965 981
FT DISULFID 983 992
FT DISULFID 995 1007
FT DISULFID 1010 1019
FT DISULFID 1012 1026
FT DISULFID 1028 1037
FT DISULFID 1040 1053
FT DISULFID 1056 1068
FT DISULFID 1058 1075
FT DISULFID 1077 1086
FT DISULFID 1089 1099
FT DISULFID 1416 1425
FT DISULFID 1418 1432
FT DISULFID 1435 1444
FT DISULFID 1447 1462
FT DISULFID 1465 1480
FT DISULFID 1467 1490
FT DISULFID 1493 1502
FT DISULFID 1505 1520
FT DISULFID 1523 1535
FT DISULFID 1525 1542
FT DISULFID 1544 1553
FT DISULFID 1556 1567
FT DISULFID 1570 1574
FT DISULFID 1574 1574
FT CARBOHYD 51 51
FT CARBOHYD 85 85
FT CARBOHYD 299 299
FT CARBOHYD 359 359
FT CARBOHYD 376 376
FT CARBOHYD 466 466
FT CARBOHYD 742 742
FT CARBOHYD 919 919
FT CARBOHYD 1031 1031
FT CARBOHYD 1057 1057
FT CARBOHYD 1593 1593
FT CARBOHYD 1610 1610
FT CARBOHYD 1696 1696
FT CARBOHYD 1806 1806
FT CARBOHYD 1897 1897
FT CARBOHYD 1912 1912
Query Match 73.3%; Score 44; DB 1; Length 3106;
Best Local Similarity 70.0%; Pred. No. 5.8;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 1 CVIGYSGDRC 10
Db 831 CPVIGTGPC 840
RESULT 11
ID LML2_CAEL STANDARD; PRT; 3672 AA.
AC Q21313;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE LAMININ-LIKE PROTEIN K08C7.3 PRECURSOR.
GN K08C7.3.
OS Caenorhabditis elegans.

```

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Berks M.;  
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).  
 CC -!- SIMILARITY: CONTAINS 21-5 LAMININ EGF-LIKE DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Z70286; CAA94293.1; -;  
 CC HSPP; P02468; IKLO.  
 CC WormPep; K08C7.3; CE06136.  
 CC InterPro; IPR000034; -;  
 CC InterPro; IPR000561; -;  
 CC InterPro; IPR001791; -;  
 CC InterPro; IPR001886; -;  
 CC InterPro; IPR002049; -;  
 CC Pfam; PF00052; laminin\_B; 1.  
 CC Pfam; PF00053; laminin\_EGF; 21.  
 CC Pfam; PF00054; laminin\_G; 5.  
 CC Pfam; PF00055; laminin\_Nterm; 1.  
 CC PRINTS; PR00011; EGF\_LAMININ.  
 CC PROSITE; PS00022; EGF\_1; 19.  
 CC PROSITE; PS01186; EGF\_2; 4.  
 CC PROSITE; PS01248; LAMININ\_TYPE\_EGF; 21.  
 KW Hypothetical protein; Laminin EGF-like domain; Signal; Repeat.  
 FT SIGNAL 1 27 POTENTIAL.  
 FT CHAIN 28 3672 LAMININ-LIKE PROTEIN K08C7.3.  
 FT DOMAIN 28 297 LAMININ N-TERMINAL (DOMAIN VI).  
 FT DOMAIN 298 356 LAMININ EGF-LIKE 1.  
 FT DOMAIN 357 426 LAMININ EGF-LIKE 2.  
 FT DOMAIN 427 471 LAMININ EGF-LIKE 3.  
 FT DOMAIN 472 518 LAMININ EGF-LIKE 4.  
 FT DOMAIN 519 563 LAMININ EGF-LIKE 5.  
 FT DOMAIN 564 609 LAMININ EGF-LIKE 6.  
 FT DOMAIN 610 655 LAMININ EGF-LIKE 7.  
 FT DOMAIN 656 700 LAMININ EGF-LIKE 8.  
 FT DOMAIN 701 755 LAMININ EGF-LIKE 9.  
 FT DOMAIN 756 808 LAMININ EGF-LIKE 10.  
 FT DOMAIN 809 839 LAMININ EGF-LIKE 11 (INCOMPLETE).  
 FT DOMAIN 1415 1460 LAMININ EGF-LIKE 12.  
 FT DOMAIN 1461 1505 LAMININ EGF-LIKE 13.  
 FT DOMAIN 1506 1553 LAMININ EGF-LIKE 14.  
 FT DOMAIN 1554 1604 LAMININ EGF-LIKE 15.  
 FT DOMAIN 1605 1614 LAMININ EGF-LIKE 16 (N-TERMINAL).  
 FT DOMAIN 1615 1796 LAMININ DOMAIN IV.  
 FT DOMAIN 1797 1829 LAMININ EGF-LIKE 16 (C-TERMINAL).  
 FT DOMAIN 1830 1879 LAMININ EGF-LIKE 17.  
 FT DOMAIN 1880 1936 LAMININ EGF-LIKE 18.  
 FT DOMAIN 1937 1989 LAMININ EGF-LIKE 19.  
 FT DOMAIN 1990 2036 LAMININ EGF-LIKE 20.  
 FT DOMAIN 2037 2083 LAMININ EGF-LIKE 21.  
 FT DOMAIN 2084 2131 LAMININ EGF-LIKE 22.  
 FT DOMAIN 2132 307 BY SIMILARITY.  
 FT DISULFID 300 320 BY SIMILARITY.  
 FT DISULFID 322 331 BY SIMILARITY.  
 FT DISULFID 334 354 BY SIMILARITY.  
 FT DISULFID 357 366 BY SIMILARITY.  
 FT DISULFID 359 391 BY SIMILARITY.  
 FT DISULFID 394 403 BY SIMILARITY.  
 FT DISULFID 406 424 BY SIMILARITY.  
 FT DISULFID 427 438 BY SIMILARITY.  
 FT DISULFID 439 445 BY SIMILARITY.  
 FT DISULFID 447 456 BY SIMILARITY.  
 FT DISULFID 459 469 BY SIMILARITY.  
 FT DISULFID 472 484 BY SIMILARITY.  
 FT DISULFID 474 491 BY SIMILARITY.  
 FT DISULFID 493 502 BY SIMILARITY.  
 FT DISULFID 505 516 BY SIMILARITY.  
 FT DISULFID 519 531 BY SIMILARITY.  
 FT DISULFID 531 538 BY SIMILARITY.  
 FT DISULFID 540 549 BY SIMILARITY.  
 FT DISULFID 552 561 BY SIMILARITY.  
 FT DISULFID 564 576 BY SIMILARITY.  
 FT DISULFID 566 583 BY SIMILARITY.  
 FT DISULFID 585 594 BY SIMILARITY.  
 FT DISULFID 597 607 BY SIMILARITY.  
 FT DISULFID 610 622 BY SIMILARITY.  
 FT DISULFID 612 629 BY SIMILARITY.  
 FT DISULFID 631 640 BY SIMILARITY.  
 FT DISULFID 643 653 BY SIMILARITY.  
 FT DISULFID 656 668 BY SIMILARITY.  
 FT DISULFID 658 674 BY SIMILARITY.  
 FT DISULFID 676 685 BY SIMILARITY.  
 FT DISULFID 688 698 BY SIMILARITY.  
 FT DISULFID 701 715 BY SIMILARITY.  
 FT DISULFID 703 724 BY SIMILARITY.  
 FT DISULFID 726 735 BY SIMILARITY.  
 FT DISULFID 738 753 BY SIMILARITY.  
 FT DISULFID 756 770 BY SIMILARITY.  
 FT DISULFID 758 777 BY SIMILARITY.  
 FT DISULFID 779 788 BY SIMILARITY.  
 FT DISULFID 791 806 BY SIMILARITY.  
 FT DISULFID 1415 1427 BY SIMILARITY.  
 FT DISULFID 1417 1434 BY SIMILARITY.  
 FT DISULFID 1436 1445 BY SIMILARITY.  
 FT DISULFID 1448 1458 BY SIMILARITY.  
 FT DISULFID 1461 1469 BY SIMILARITY.  
 FT DISULFID 1463 1476 BY SIMILARITY.  
 FT DISULFID 1478 1487 BY SIMILARITY.  
 FT DISULFID 1490 1503 BY SIMILARITY.  
 FT DISULFID 1506 1520 BY SIMILARITY.  
 FT DISULFID 1508 1527 BY SIMILARITY.  
 FT DISULFID 1529 1538 BY SIMILARITY.  
 FT DISULFID 1541 1551 BY SIMILARITY.  
 FT DISULFID 1554 1566 BY SIMILARITY.  
 FT DISULFID 1556 1573 BY SIMILARITY.  
 FT DISULFID 1575 1584 BY SIMILARITY.  
 FT DISULFID 1587 1602 BY SIMILARITY.  
 FT DISULFID 1630 1639 BY SIMILARITY.  
 FT DISULFID 1632 1646 BY SIMILARITY.  
 FT DISULFID 1649 1658 BY SIMILARITY.  
 FT DISULFID 1661 1677 BY SIMILARITY.  
 FT DISULFID 1680 1694 BY SIMILARITY.  
 FT DISULFID 1682 1695 BY SIMILARITY.  
 FT DISULFID 1697 1706 BY SIMILARITY.  
 FT DISULFID 1719 1734 BY SIMILARITY.  
 FT DISULFID 1737 1751 BY SIMILARITY.  
 FT DISULFID 1754 1769 BY SIMILARITY.  
 FT DISULFID 1761 1770 BY SIMILARITY.  
 FT DISULFID 1773 1787 BY SIMILARITY.  
 FT DISULFID 1790 1800 BY SIMILARITY.  
 FT DISULFID 1832 1846 BY SIMILARITY.  
 FT DISULFID 1849 1858 BY SIMILARITY.  
 FT DISULFID 1861 1877 BY SIMILARITY.  
 FT DISULFID 1880 1894 BY SIMILARITY.  
 FT DISULFID 1882 1905 BY SIMILARITY.  
 FT DISULFID 1907 1916 BY SIMILARITY.  
 FT DISULFID 1919 1934 BY SIMILARITY.  
 FT DISULFID 1937 1951 BY SIMILARITY.  
 FT DISULFID 1939 1958 BY SIMILARITY.  
 FT DISULFID 1961 1970 BY SIMILARITY.  
 FT DISULFID 1973 1987 BY SIMILARITY.  
 FT DISULFID 1990 2000 BY SIMILARITY.  
 FT DISULFID 1992 2007 BY SIMILARITY.  
 FT DISULFID 2009 2018 BY SIMILARITY.  
 FT DISULFID 2021 2034 BY SIMILARITY.  
 FT DISULFID 2037 2048 BY SIMILARITY.  
 FT DISULFID 2039 2055 BY SIMILARITY.  
 FT DISULFID 2057 2066 BY SIMILARITY.  
 FT DISULFID 2069 2081 BY SIMILARITY.  
 FT DISULFID 2084 2096 BY SIMILARITY.  
 FT DISULFID 2086 2103 BY SIMILARITY.  
 FT DISULFID 2105 2114 BY SIMILARITY.  
 FT DISULFID 2117 2129 BY SIMILARITY.  
 FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 140 140 N-LINKED (GLCNAC. . .) (POTENTIAL).



```

FT CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 511 511 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 530 530 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 634 634 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 761 761 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1014 1014 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1341 1341 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1705 1705 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1756 1756 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1868 1868 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1944 1944 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1986 1986 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2002 2002 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2159 2159 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2207 2207 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2231 2231 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2235 2235 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2401 2401 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2421 2421 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2487 2487 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2821 2821 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3087 3087 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3242 3242 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3541 3541 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3672 AA; 404223 MW; 28E262DB5FF14BFA CRC64;

Query Match 73.38; Score 44; DB 1; Length 3672;
Best Local Similarity 70.08; Pred. No. 6.8;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10
   I II:IIII
Db 1507 CKPGYTGDCR 1916

RESULT 12
FBPI_STRPU ID FBPI_STRPU STANDARD; PRT; 1064 AA.
AC P10079;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE FBROPELLIN I PRECURSOR (EPIDERMAL GROWTH FACTOR-RELATED PROTEIN 1)
DE (UEGF-1).
GN EGF1.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN 11;
RP SEQUENCE FROM N.A.
RX MEDLINE=90112459; PubMed=2514273;
RA Delgadillo-Reynoso M.G., Rolio D.R., Hursh D.A., Raff R.A.;
RT "Structural analysis of the uEGF gene in the sea urchin
RT strongylocentrotus purpuratus reveals more similarity to vertebrate
RT than to invertebrate genes with EGF-like repeats.";
RL J. Mol. Evol. 29:314-327(1989).
RN [2]
RP SEQUENCE OF 279-476 AND 781-1064 FROM N.A.
RX MEDLINE=87319677; PubMed=3498216;
RA Hursh D.A., Andrews M.E., Raff R.A.;
RT "A sea urchin gene encodes a polypeptide homologous to epidermal
RT growth factor.";
RL Science 237:1487-1490(1987).
RN [3]
RP AVIDIN-LIKE DOMAIN.
RX MEDLINE=89196806; PubMed=2784773;
RA Hunt L.T., Barker W.C.;
RT "Avidin-like domain in an epidermal growth factor homolog from a sea

```

```

RT urchin.";
RL FASEB J. 3:1760-1764(1989).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=91285254; PubMed=2060714;
RA Bisgrove B.W., Andrews M.E., Raff R.A.;
RT "Fibropellins, products of an EGF repeat-containing gene, form a
RT unique extracellular matrix structure that surrounds the sea urchin
RT embryo.";
RL Dev. Biol. 146:89-99(1991).
CC -!- FUNCTION: FORM THE APICAL LAMINA, A COMPONENT OF THE EXTRACELLULAR
CC MATRIX.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR. IN VESICLES IN THE CYTOPLASM
CC OF UNFERTILIZED EGGS, THEN TO THE BASE OF THE HYALIN LAYER
CC THROUGHOUT DEVELOPMENT AND FINALLY IN THE APICAL LAMINA IN LATE
CC EMBRYOS AND EARLY LARVAE.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; IA (SHOWN HERE) AND IB; ARE
CC PRODUCED BY ALTERNATIVE SPLICING. THE SMALL FORM (IB) LACKS 8 EGF
CC REPEATS.
CC -!- DEVELOPMENTAL STAGE: MODERATE LEVELS IN UNFERTILIZED EGGS AND
CC DURING EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN
CC LATE MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS
CC MAINTAINED THROUGH SUBSEQUENT STAGES. EXPRESSED BOTH MATERNALLY
CC AND ZYGOTICALLY.
CC -!- SIMILARITY: CONTAINS 21 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC -!- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR
CC TO AVIDIN/STREPTAVIDIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L08692; AAA62164.1; -
DR EMBL; L08692; AAA62163.1; -
DR EMBL; X17530; CAA35571.1; -
DR EMBL; X17421; AAA30050.1; -
DR EMBL; X17533; CAA35573.1; -
DR PIR; A29316; A29316.
DR HSSP; P01132; IEPH.
DR InterPro; IPR000088; -
DR InterPro; IPR000152; -
DR InterPro; IPR000561; -
DR InterPro; IPR000859; -
DR InterPro; IPR001438; -
DR InterPro; IPR001881; -
DR Pfam; PF01382; Avidin; 1.
DR Pfam; PF00431; CUB; 1.
DR Pfam; PF00008; EGF; 21.
DR PRINTS; PR00010; EGFBL00D.
DR PROSITE; PS00010; ASX_HYDROXYL; 19.
DR PROSITE; PS00022; EGF_1; 19.
DR PROSITE; PS00577; AVIDIN; 1.
DR PROSITE; PS01180; CUB; 1.
DR PROSITE; PS01186; EGF_2; 19.
DR PROSITE; PS01187; EGF_CA; 19.
KW Biotin; Alternative splicing; EGF-like domain; Repeat; Signal;
KW Glycoprotein; Calcium-binding.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1064 FIBROPELLIN I.
FT DOMAIN 20 55 EGF-LIKE 1.
FT DOMAIN 52 175 CUB.
FT DOMAIN 176 212 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 214 250 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 252 288 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 290 326 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 328 364 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 366 402 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 404 440 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).

```

FT DOMAIN 442 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 480 EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 516 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 554 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 592 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 630 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 668 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 706 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 744 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 782 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 820 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 858 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 896 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).  
FT DOMAIN 934 AVIDIN-LIKE.  
FT DISULFID 936 BY SIMILARITY.  
FT DISULFID 23 BY SIMILARITY.  
FT DISULFID 28 BY SIMILARITY.  
FT DISULFID 45 BY SIMILARITY.  
FT DISULFID 180 BY SIMILARITY.  
FT DISULFID 185 BY SIMILARITY.  
FT DISULFID 202 BY SIMILARITY.  
FT DISULFID 218 BY SIMILARITY.  
FT DISULFID 223 BY SIMILARITY.  
FT DISULFID 238 BY SIMILARITY.  
FT DISULFID 240 BY SIMILARITY.  
FT DISULFID 249 BY SIMILARITY.  
FT DISULFID 256 BY SIMILARITY.  
FT DISULFID 261 BY SIMILARITY.  
FT DISULFID 276 BY SIMILARITY.  
FT DISULFID 287 BY SIMILARITY.  
FT DISULFID 294 BY SIMILARITY.  
FT DISULFID 299 BY SIMILARITY.  
FT DISULFID 316 BY SIMILARITY.  
FT DISULFID 332 BY SIMILARITY.  
FT DISULFID 337 BY SIMILARITY.  
FT DISULFID 354 BY SIMILARITY.  
FT DISULFID 354 BY SIMILARITY.  
FT DISULFID 370 BY SIMILARITY.  
FT DISULFID 375 BY SIMILARITY.  
FT DISULFID 392 BY SIMILARITY.  
FT DISULFID 401 BY SIMILARITY.  
FT DISULFID 408 BY SIMILARITY.  
FT DISULFID 413 BY SIMILARITY.  
FT DISULFID 428 BY SIMILARITY.  
FT DISULFID 430 BY SIMILARITY.  
FT DISULFID 439 BY SIMILARITY.  
FT DISULFID 446 BY SIMILARITY.  
FT DISULFID 451 BY SIMILARITY.  
FT DISULFID 468 BY SIMILARITY.  
FT DISULFID 484 BY SIMILARITY.  
FT DISULFID 489 BY SIMILARITY.  
FT DISULFID 504 BY SIMILARITY.  
FT DISULFID 506 BY SIMILARITY.  
FT DISULFID 515 BY SIMILARITY.  
FT DISULFID 522 BY SIMILARITY.  
FT DISULFID 527 BY SIMILARITY.  
FT DISULFID 544 BY SIMILARITY.  
FT DISULFID 560 BY SIMILARITY.  
FT DISULFID 565 BY SIMILARITY.  
FT DISULFID 582 BY SIMILARITY.  
FT DISULFID 598 BY SIMILARITY.  
FT DISULFID 603 BY SIMILARITY.  
FT DISULFID 620 BY SIMILARITY.  
FT DISULFID 636 BY SIMILARITY.  
FT DISULFID 641 BY SIMILARITY.  
FT DISULFID 658 BY SIMILARITY.  
FT DISULFID 674 BY SIMILARITY.  
FT DISULFID 679 BY SIMILARITY.  
FT DISULFID 696 BY SIMILARITY.  
FT DISULFID 712 BY SIMILARITY.  
FT DISULFID 717 BY SIMILARITY.  
FT DISULFID 734 BY SIMILARITY.  
FT DISULFID 743 BY SIMILARITY.  
FT DISULFID 750 BY SIMILARITY.  
FT DISULFID 755 BY SIMILARITY.  
FT DISULFID 772 BY SIMILARITY.  
FT DISULFID 788 BY SIMILARITY.  
FT DISULFID 793 BY SIMILARITY.  
FT DISULFID 810 BY SIMILARITY.  
FT DISULFID 826 BY SIMILARITY.  
FT DISULFID 831 BY SIMILARITY.  
FT DISULFID 846 BY SIMILARITY.  
FT DISULFID 857 BY SIMILARITY.  
FT DISULFID 864 BY SIMILARITY.  
FT DISULFID 875 BY SIMILARITY.  
FT DISULFID 884 BY SIMILARITY.

FT DISULFID 886 895 BY SIMILARITY.  
FT DISULFID 902 913 BY SIMILARITY.  
FT DISULFID 907 922 BY SIMILARITY.  
FT DISULFID 924 933 BY SIMILARITY.  
FT CARBOHYD 30 30 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 851 851 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARSPLIC 477 780 MISSING (IN ISOFORM IB).  
FT CONFLICT 279 279 L -> S (IN REF. 2).  
SQ SEQUENCE 1064 AA; 112072 MW; 2E569CA012ED6D09 CRC64;  
  
Query Match 71.7%; Score 43; DB 1; Length 1064;  
Best Local Similarity 60.0%; Pred. No. 3.1;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 CVIGYSGDRC 10  
DB 316 CPLFGSGDNC 325  
| : : : : : |  
| : : : : : |  
  
RESULT 13  
CADN\_DROME  
ID CADN\_DROME STANDARD; PRT; 3097 AA.  
AC 015943; QGVJB7;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE NEURAL-CADHERIN PRECURSOR (CADHERIN-N PROTEIN) (DN-CADHERIN).  
GN CADN OR CG7100.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Head, and Embryo;  
RX MEDLINE=9738431; PubMed=9247265;  
RA Iwai Y., Usui T., Hirano S., Steward R., Takeichi M., Uemura T.;  
RT "Axon patterning requires DN-cadherin, a novel neuronal adhesion  
receptor, in the Drosophila embryonic CNS.";  
RL Neuron 19:77-89(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers J.-H.C., Blazer E.G., Helt R.G., Champagne M., Pfeiffer B.D.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cavener L.B., Davies P.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Dew I., Dietz S.M.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Wei M.-H., Ibegwam C.,  
RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,  
RA Foshier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Sidman Klamis I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svrlkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*";  
RL Science 287:2185-2195(2000).  
RN [3]  
RP INTERACTION WITH ARM.  
RX MEDLINE=98298928; PubMed=9635189;  
RA Loureiro J., Peifer M.;  
RA "Roles of Armadillo, a *Drosophila* catenin, during central nervous  
RT system development";  
RT Curr. Biol. 8:622-632(1998).  
RL  
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE  
CC SORTING OF HETEROGENEOUS CELL TYPES. MAY ASSOCIATE WITH ARM NEURAL  
CC ISOFORM AND PARTICIPATE IN THE TRANSMISSION OF DEVELOPMENTAL  
CC INFORMATION.  
CC  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).  
CC  
CC -!- TISSUE SPECIFICITY: IN THE EMBRYO, THE PROTEIN FIRST APPEARS IN  
CC THE MESODERM AT STAGE 9 AND IS PRESENT IN THE MYOBLASTS AND MUSCLE  
CC FIBERS BY STAGE 12 AND STAGE 14, RESPECTIVELY. AT STAGE 12 THE  
CC PROTEIN IS ALSO LOCATED IN THE AXONS OF THE ENTIRE CNS, BUT NOT IN  
CC THE GLIAL CELLS. IN THIRD INSTAR LARVAE PROTEIN IS EXPRESSED IN  
CC THE CNS NEUROPILE, PHOTORECEPTOR AXONS AND PRECURSORS OF ADULT  
CC MUSCLES.  
CC  
CC -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.  
CC  
CC -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.  
CC  
CC -!- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL; AB002397; BAA22151.1; -;  
CC DR EMBL; AE003656; AAF53635.1; -;  
CC DR HSP; P00740; 11XA  
CC DR FlyBase; FBgn0015609; Cadn.  
CC DR InterPro; IPR000233; -;  
CC DR InterPro; IPR000561; -;  
CC DR InterPro; IPR000742; -;  
CC DR InterPro; IPR001791; -;  
CC DR InterPro; IPR002126; -;  
CC DR Pfam; PF00008; EGF; 3.  
CC DR Pfam; PF00028; cadherin; 14.  
CC DR Pfam; PF00054; laminin\_G; 2.  
CC DR Pfam; PF01049; Cadherin\_C term; 1.  
CC DR PRINTS; PR00205; CADHERIN.  
CC DR PROSITE; PS00232; CADHERIN; 9.  
CC DR PROSITE; PS00022; EGF\_1; 3.  
CC DR PROSITE; PS01186; EGF\_2; 3.  
CC KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;  
CC KW Signal; EGF-like domain.  
CC FT SIGNAL 1 36  
CC FT PROPEP 37 ?  
CC FT CHAIN ? 3097  
CC FT DOMAIN ? 1454  
CC FT TRANSMEM 1455 1475  
CC FT DOMAIN 1476 3097  
CC FT DOMAIN 181 305  
CC FT DOMAIN 430 543  
CC FT DOMAIN 554 651

FT DOMAIN 660 756  
FT DOMAIN 766 858  
FT DOMAIN 867 968  
FT DOMAIN 1078 1183  
FT DOMAIN 1087 1193  
FT DOMAIN 1193 1299  
FT DOMAIN 1307 1414  
FT DOMAIN 1423 1514  
FT DOMAIN 1523 1630  
FT DOMAIN 1639 1742  
FT DOMAIN 1749 1861  
FT DOMAIN 1870 1966  
FT DOMAIN 1974 2085  
FT DOMAIN 2346 2377  
FT DOMAIN 2407 2585  
FT DOMAIN 2592 2627  
FT DOMAIN 2869 2902  
FT DOMAIN 2661 2822  
FT DISULFID 2346 2357  
FT DISULFID 2351 2366  
FT DISULFID 2368 2377  
FT DISULFID 2592 2607  
FT DISULFID 2601 2616  
FT DISULFID 2618 2627  
FT DISULFID 2869 2880  
FT DISULFID 2874 2891  
FT DISULFID 2893 2902  
FT CARBOHYD 97 97  
FT CARBOHYD 150 150  
FT CARBOHYD 325 325  
FT CARBOHYD 426 426  
FT CARBOHYD 930 930  
FT CARBOHYD 1266 1266  
FT VARIANT 1425 1425  
FT  
FT CONFLICT 1342 1342  
FT CONFLICT 2786 2786  
SQ SEQUENCE 3097 AA; 347201 MW; 082242F28D9B5CC3 CRC64;  
Query Match 71.7%; Score 43; DB 1; Length 3097;  
Best Local Similarity 60.0%; Pred. No. 8.8;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CVIGYSGDRC 10  
I : I : I : I :  
Db 2368 CPVGYTGPRC 2377  
RESULT 14  
LMA2\_HUMAN  
ID LMA2\_HUMAN STANDARD; PRT; 3110 AA.  
AC P24043; Q14736;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE LAMININ ALPHA-2 CHAIN PRECURSOR (LAMININ M CHAIN) (MEROSIN HEAVY  
DE CHAIN).  
GN LAMA2 OR LAMM.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=94124633; PubMed=8294519;  
RA Vuolteenaho R., Nissinen M., Sainio K., Byers M., Eddy R.,  
RA Hirvonen H., Shows T.B., Sariola H., Engvall E., Tryggvason K.;  
RT "Human laminin M chain (merosin): complete primary structure,  
RT chromosomal assignment, and expression of the M and A chain in human  
RL fetal tissues.";  
RL J. Cell Biol. 124:381-394(1994).

[2]  
 RN SEQUENCE OF 1981-3110 FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Placenta;  
 RX MEDLINE=90238994; PubMed=2185464;  
 RA Ehrig K., Leivo I., Argraves W.S., Ruoslahti E., Engvall E.;  
 RT "Merosin, a tissue-specific basement membrane protein, is a  
 RL laminin-like protein.";   
 Proc. Natl. Acad. Sci. U.S.A. 87:3264-3268(1990).  
 [3]  
 RN VARIANTS GLN-545; HIS-619; LEU-919; HIS-2586 AND LYS-2614.  
 RA Panicker S.G., Mendell J.T., Chen L., Feng B., Sahenk Z.,  
 RA Marzluf G.A., Amato A.A., Mendell J.R.;  
 RT "Novel single base polymorphisms and rare sequence variants in  
 RL the laminin 2-chain coding region detected by RNA/SSCP analysis.";   
 Hum. Mutat. 13:174-174(1999).  
 [4]  
 RN ERRATUM.  
 RA Panicker S.G., Mendell J.T., Chen L., Feng B., Sahenk Z.,  
 RA Marzluf G.A., Amato A.A., Mendell J.R.;  
 RL Hum. Mutat. 13:340-340(1999).  
 CC -!- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR. LAMININ  
 CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF  
 CC CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING  
 CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.  
 CC -!- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE  
 CC DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND  
 CC TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE  
 CC COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.  
 CC THE ALPHA-2 CHAIN IS A SUBUNIT OF LAMININ-2 (MEROSIN) AND LAMININ-  
 CC 4 (S-MEROSIN).  
 CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT  
 CC MEMBRANES (MAJOR COMPONENT).  
 CC -!- TISSUE SPECIFICITY: PLACENTA, STRIATED MUSCLE, PERIPHERAL NERVE,  
 CC CARDIAC MUSCLE, PANCREAS, LUNG, SPLEEN, KIDNEY, ADRENAL GLAND,  
 CC SKIN, TESTIS, MENINGES, CHOROID PLEXUS, AND SOME OTHER REGIONS OF  
 CC THE BRAIN; NOT IN LIVER, THYMUS AND BONE.  
 CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT  
 CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.  
 CC -!- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.  
 CC -!- DISEASE: DEFECTS IN LAMA2 ARE THE CAUSE OF MEROSIN-DEFICIENT  
 CC CONGENITAL MUSCULAR DYSTROPHY (MCMD).  
 CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).  
 CC -!- SIMILARITY: CONTAINS 17 LAMININ EGF-LIKE DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.  
 CC -!- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z26653; CAA81394.1; -;  
 DR EMBL; M59832; AAG3215.1; -;  
 DR PIR; A35899; MMHUMH.  
 DR HSP; P02468; 1KLO.  
 DR MIM; 156225; -;  
 DR InterPro; IPR000034; -;  
 DR InterPro; IPR000561; -;  
 DR InterPro; IPR001791; -;  
 DR InterPro; IPR001886; -;  
 DR InterPro; IPR002049; -;  
 DR Pfam; PF00052; laminin\_B; 2;  
 DR Pfam; PF00053; laminin\_EGF; 15;  
 DR Pfam; PF00054; laminin\_G; 5;  
 DR Pfam; PF00055; laminin\_Nterm; 1;  
 DR PRINTS; PR00011; EGF\_LAMININ.  
 DR PROSITE; PS00022; EGF\_1; 11;  
 DR PROSITE; PS01186; EGF\_2; 3;  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 14;  
 KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;

KW	Laminin	EGF-like	domain;	Cell	adhesion;	Repeat;	Signal;	Polymorphism.
FT	SIGNAL	1	22		POTENTIAL.			
FT	CHAIN	23	3110		LAMININ ALPHA-2 CHAIN.			
FT	DOMAIN	23	286		LAMININ N-TERMINAL (DOMAIN VI).			
FT	DOMAIN	287	527		4.5 X LAMININ EGF-LIKE REPEATS.			
FT	DOMAIN	287	343		LAMININ EGF-LIKE 1.			
FT	DOMAIN	344	413		LAMININ EGF-LIKE 2.			
FT	DOMAIN	414	468		LAMININ EGF-LIKE 3.			
FT	DOMAIN	469	517		LAMININ EGF-LIKE 4.			
FT	DOMAIN	518	527		LAMININ EGF-LIKE 5 (N-TERMINAL).			
FT	DOMAIN	528	723		LAMININ DOMAIN IV 1 (DOMAIN IV B).			
FT	DOMAIN	724	1175		9 X LAMININ EGF-LIKE REPEATS (DOMAIN III B).			
FT	DOMAIN	724	756		LAMININ EGF-LIKE 5 (C-TERMINAL).			
FT	DOMAIN	757	806		LAMININ EGF-LIKE 6.			
FT	DOMAIN	807	864		LAMININ EGF-LIKE 7.			
FT	DOMAIN	865	917		LAMININ EGF-LIKE 8.			
FT	DOMAIN	918	966		LAMININ EGF-LIKE 9.			
FT	DOMAIN	967	1013		LAMININ EGF-LIKE 10.			
FT	DOMAIN	1014	1059		LAMININ EGF-LIKE 11.			
FT	DOMAIN	1060	1105		LAMININ EGF-LIKE 12.			
FT	DOMAIN	1106	1165		LAMININ EGF-LIKE 13.			
FT	DOMAIN	1166	1175		LAMININ EGF-LIKE 14 (N-TERMINAL).			
FT	DOMAIN	1176	1379		LAMININ DOMAIN IV 2 (DOMAIN IV A).			
FT	DOMAIN	1380	1573		3.5 X LAMININ EGF-LIKE REPEATS.			
FT	DOMAIN	1380	1419		LAMININ EGF-LIKE 14 (C-TERMINAL).			
FT	DOMAIN	1420	1468		LAMININ EGF-LIKE 15.			
FT	DOMAIN	1469	1526		LAMININ EGF-LIKE 16.			
FT	DOMAIN	1527	1573		LAMININ EGF-LIKE 17.			
FT	DOMAIN	1581	2154		DOMAIN II AND I.			
FT	DOMAIN	2155	3110		5 X LAMININ G-LIKE REPEATS (DOMAIN G).			
FT	DOMAIN	2169	2363		LAMININ G-LIKE 1.			
FT	DOMAIN	2364	2550		LAMININ G-LIKE 2.			
FT	DOMAIN	2551	2787		LAMININ G-LIKE 3.			
FT	DOMAIN	2788	2962		LAMININ G-LIKE 4.			
FT	DOMAIN	2963	3110		LAMININ G-LIKE 5.			
FT	DOMAIN	1630	2150		COILED COIL (POTENTIAL).			
FT	DISULFID	287	296		BY SIMILARITY.			
FT	DISULFID	289	307		BY SIMILARITY.			
FT	DISULFID	309	318		BY SIMILARITY.			
FT	DISULFID	321	341		BY SIMILARITY.			
FT	DISULFID	344	353		BY SIMILARITY.			
FT	DISULFID	346	378		BY SIMILARITY.			
FT	DISULFID	381	390		BY SIMILARITY.			
FT	DISULFID	393	411		BY SIMILARITY.			
FT	DISULFID	414	426		BY SIMILARITY.			
FT	DISULFID	416	442		BY SIMILARITY.			
FT	DISULFID	444	453		BY SIMILARITY.			
FT	DISULFID	456	466		BY SIMILARITY.			
FT	DISULFID	469	482		BY SIMILARITY.			
FT	DISULFID	471	486		BY SIMILARITY.			
FT	DISULFID	488	497		BY SIMILARITY.			
FT	DISULFID	500	515		BY SIMILARITY.			
FT	DISULFID	507	515		BY SIMILARITY.			
FT	DISULFID	759	773		BY SIMILARITY.			
FT	DISULFID	776	785		BY SIMILARITY.			
FT	DISULFID	788	804		BY SIMILARITY.			
FT	DISULFID	807	822		BY SIMILARITY.			
FT	DISULFID	809	832		BY SIMILARITY.			
FT	DISULFID	835	844		BY SIMILARITY.			
FT	DISULFID	847	862		BY SIMILARITY.			
FT	DISULFID	865	879		BY SIMILARITY.			
FT	DISULFID	889	898		BY SIMILARITY.			
FT	DISULFID	901	915		BY SIMILARITY.			
FT	DISULFID	918	930		BY SIMILARITY.			
FT	DISULFID	920	937		BY SIMILARITY.			
FT	DISULFID	939	948		BY SIMILARITY.			
FT	DISULFID	951	964		BY SIMILARITY.			
FT	DISULFID	967	979		BY SIMILARITY.			
FT	DISULFID	969	985		BY SIMILARITY.			
FT	DISULFID	987	996		BY SIMILARITY.			
FT	DISULFID	999	1011		BY SIMILARITY.			

```

FT DISULFID 1014 1023 BY SIMILARITY.
FT DISULFID 1016 1030 BY SIMILARITY.
FT DISULFID 1032 1041 BY SIMILARITY.
FT DISULFID 1044 1057 BY SIMILARITY.
FT DISULFID 1060 1072 BY SIMILARITY.
FT DISULFID 1062 1079 BY SIMILARITY.
FT DISULFID 1081 1090 BY SIMILARITY.
FT DISULFID 1093 1103 BY SIMILARITY.
FT DISULFID 1420 1429 BY SIMILARITY.
FT DISULFID 1422 1436 BY SIMILARITY.
FT DISULFID 1439 1448 BY SIMILARITY.
FT DISULFID 1451 1466 BY SIMILARITY.
FT DISULFID 1469 1484 BY SIMILARITY.
FT DISULFID 1471 1494 BY SIMILARITY.
FT DISULFID 1497 1506 BY SIMILARITY.
FT DISULFID 1509 1524 BY SIMILARITY.
FT DISULFID 1527 1539 BY SIMILARITY.
FT DISULFID 1529 1546 BY SIMILARITY.
FT DISULFID 1548 1557 BY SIMILARITY.
FT DISULFID 1560 1571 BY SIMILARITY.
FT DISULFID 1574 1578 INTERCHAIN (PROBABLE).
FT CARBOHYD 55 55 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 303 303 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 363 363 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 380 380 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 746 746 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1061 1061 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1597 1597 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1614 1614 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 71.7%; Score 43; DB 1; Length 3110;
Best Local Similarity 60.0%; Pred. No. 8.8;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CVIGSGDRC 10
  | :||: ||
-Db 835 CPVGTGPRC 844

RESULT 15
DL_DROME STANDARD; PRT; 833 AA.
AC P10041: Q9VDY2; Q99108;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEUROGENIC LOCUS DELTA PROTEIN PRECURSOR.
GN DL OR CG3619.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
RC TISSUE=Embryo;
RA Vaessin H., Bremer K.A., Knust E., Campos-Ortega J.A.;
RT "The neurogenic gene Delta of Drosophila melanogaster is expressed in neurogenic territories and encodes a putative transmembrane protein with EGF-like repeats.";
RL EMBO J. 6:3431-3440(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R; TISSUE=Embryo;
RX MEDLINE=89196890; PubMed=3149249;
RA Koczynski C.C., Alton A.K., Fechtel K., Kooh P.J., Muskavitch M.A.T.;
RT "Delta, a Drosophila neurogenic gene, is transcriptionally complex and encodes a protein related to blood coagulation factors and epidermal growth factor of vertebrates.";
RL Genes Dev. 2:1723-1735(1988).

```

```

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Cocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbavani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Patuzzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [4]
RP SEQUENCE OF 422-621 FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=87218537; PubMed=3107986;
RA Knust E., Dietrich U., Tepass U., Bremer K.A., Weigel D., Vaessin H.,
RA Campos-Ortega J.A.;
RT "EGF homologous sequences encoded in the genome of Drosophila melanogaster, and their relation to neurogenic genes.";
RL EMBO J. 6:761-766(1987).
RN [5]
RP PATTERN OF TRANSCRIPTION, AND CHARACTERIZATION.
RX MEDLINE=91209246; PubMed=2128477;
RA Haenlin M., Kramatschek B., Campos-Ortega J.A.;
RT "The pattern of transcription of the neurogenic gene Delta of Drosophila melanogaster.";
RL Development 110:905-914(1990).
CC -!- FUNCTION: ESSENTIAL FOR PROPER DIFFERENTIATION OF ECTODERM. DL IS REQUIRED FOR THE CORRECT SEPARATION OF NEURAL AND EPIDERMAL CELL LINEAGES.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: DETECTED IN ALL AREAS WITH NEUROGENIC ABILITIES. FOR EXAMPLE THE NEUROGENIC ECTODERM AND THE PRIMITIVE OF THE SENSE ORGANS. LATER EXPRESSION IS RESTRICTED TO THOSE CELLS THAT HAVE ADOPTED A NEURAL FATE.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED BOTH MATERNALLY AND ZYGOTICALLY. EXPRESSION IS HIGHEST EARLY IN EMBRYONIC DEVELOPMENT (STAGE 5) AND REDUCES TO A LOW LEVEL DURING LARVAL STAGES.
CC -!- MISCELLANEOUS: SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO THE INNER PART OF EMBRYO IS ONE OF THE FIRST STEPS OF CNS DEVELOPMENT IN INSECTS, THIS PROCESS IS UNDER CONTROL OF THE NEUROGENIC GENES.

```

Search completed: August 15, 2001, 10:55:54  
Job time: 266 sec



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2001, 10:51:27 ; Search time 33.36 Seconds  
(without alignments)  
39.660 Million cell updates/sec

Title: US-09-673-785A-2

Perfect score: 60

Sequence: 1 CVIGSGDRC 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_16:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_unclassified:\*
- 13: sp\_vertebrate:\*
- 14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	55	91.7	2146	5	Q9VC97	Q9vc97 drosophila
2	53	88.3	656	6	O77612	O77612 bos taurus
3	53	88.3	1643	6	O77611	O77611 bos taurus
4	53	88.3	2394	6	O77610	O77610 bos taurus
5	53	88.3	3381	6	O77609	O77609 bos taurus
6	50	83.3	372	5	O21756	O21756 caenorhabdi
7	49	81.7	89	6	O28867	O28867 equus cabal
8	48	80.0	2352	5	O61240	O61240 halocynthia
9	48	80.0	4006	11	O35452	O35452 mus musculus
10	48	80.0	4114	11	O54796	O54796 mus musculus
11	47	78.3	1531	11	O9WVB5	O9wvb5 mus musculus
12	47	78.3	2920	11	O9ROM0	O9rom0 mus musculus
13	46	76.7	57	1	O29845	O29845 archaeoglob
14	46	76.7	106	4	Q9UN93	Q9un93 homo sapien
15	46	76.7	308	4	Q9UN94	Q9un94 homo sapien
16	46	76.7	366	4	Q9UN95	Q9un95 homo sapien
17	46	76.7	877	4	O9H306	O9h3q6 homo sapien
18	46	76.7	878	4	Q9H3Q7	Q9h3q7 homo sapien
19	46	76.7	878	4	Q9GZZ2	Q9gzz2 homo sapien

20	46	76.7	901	4	Q9H195	Q9h195 homo sapien
21	46	76.7	957	4	O14651	O14651 homo sapien
22	46	76.7	1217	4	Q9UKW9	Q9ukw9 homo sapien
23	45	75.0	655	11	O88564	O88564 rattus norv
24	45	75.0	963	5	O9GPM9	O9gpm9 caenorhabdi
25	45	75.0	1270	5	O9GPN0	O9gpn0 caenorhabdi
26	45	75.0	1530	11	Q9WUG5	Q9wug5 rattus norv
27	45	75.0	1531	11	O88279	O88279 rattus norv
28	45	75.0	2390	11	O08592	O08592 rattus norv
29	44	73.3	3704	5	P91904	P91904 caenorhabdi
30	43	71.7	264	5	O20043	O20043 caenorhabdi
31	43	71.7	379	11	O35883	O35883 rattus norv
32	43	71.7	447	11	O63348	O63348 rattus norv
33	43	71.7	711	5	O9XWC4	O9xwc4 caenorhabdi
34	43	71.7	1722	5	Q19350	Q19350 caenorhabdi
35	43	71.7	1783	5	Q9VJB6	Q9vjb6 drosophila
36	43	71.7	2319	11	O9R172	O9r172 rattus norv
37	43	71.7	2656	5	Q9GNU3	Q9gnu3 paracentrot
38	43	71.7	3110	4	Q93022	Q93022 homo sapien
39	42	70.0	403	5	O18375	O18375 drosophila
40	42	70.0	559	5	Q9VZ44	Q9vz44 drosophila
41	42	70.0	664	4	Q9UIL7	Q9uil7 homo sapien
42	42	70.0	752	13	O42374	O42374 brachydanio
43	42	70.0	1534	4	O75093	O75093 homo sapien
44	42	70.0	2281	4	Q9UPL3	Q9upl3 homo sapien
45	42	70.0	2321	4	Q9Y6L8	Q9y6l8 homo sapien

#### ALIGNMENTS

RESULT 1

ID Q9VC97 PRELIMINARY; PRT; 2146 AA.

AC Q9VC97;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)

DE CRB PROTEIN.

GN CRB OR CG6383.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson A.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M., M.C.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RA "The genome sequence of Drosophila melanogaster";  
RL Science 287:2185-2195(2000).  
DR EMBL; AE003747; AAF56276.1; -;  
DR HSSP; P00740; 1IXA.  
DR FlyBase; FBgn000368; crb.  
DR InterPro; IPR000152; -;  
DR InterPro; IPR000561; -;  
DR InterPro; IPR000742; -;  
DR InterPro; IPR001438; -;  
DR InterPro; IPR001791; -;  
DR InterPro; IPR001881; -;  
DR InterPro; IPR02049; -;  
DR Pfam; PF00008; EGF; 27.  
DR Pfam; PF00054; laminin\_G; 3.  
DR PRINTS; PR00010; EGFBLLOOD.  
DR PRINTS; PR00011; EGFAMININ.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 16.  
DR PROSITE; PS00022; EGF\_1; 25.  
DR PROSITE; PS01186; EGF\_2; 17.  
DR PROSITE; PS01187; EGF\_CA; 13.  
DR SMART; SM00179; EGF\_CA; 1.  
DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
KW SIGNAL  
FT CHAIN 21 656  
SQ SEQUENCE 2146 AA; 233570 MW; 8E23B9E32B761115 CRC64;  
  
Query Match 91.7%; Score 55; DB 5; Length 2146;  
Best Local Similarity 80.0%; Pred. No. 0.24; 1; Indels 0; Gaps 0;  
Matches 8; Conservative 1; Mismatches 0;  
  
Qy 1 CVIGYSGDRC 10  
Db 569 CAVGYSGDRC 578  
I:|||||||  
-:|||||||  
  
RESULT 2  
077612 PRELIMINARY; PRT; 656 AA.  
AC 077612;  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE VERSICAN V3 SPLICE-VARIANT PRECURSOR.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98288320; PubMed=9624174;  
RA Schmalfeldt M., Dours-Zimmermann M.T., Winterhalter K.H.,  
RT "Versican V2 is a major extracellular matrix component of the mature bovine brain."  
RL J. Biol. Chem. 273:15758-15764(1998).  
CC -|- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX DOMAIN  
DR EMBL; AF060459; AAC24361.1; -;  
DR HSSP; P01132; 1EGF.  
DR InterPro; IPR000152; -;  
DR InterPro; IPR000436; -;  
RT "Versican V2 is a major extracellular matrix component of the mature bovine brain."  
RL J. Biol. Chem. 273:15758-15764(1998).  
CC -|- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX DOMAIN  
DR EMBL; AF060459; AAC24361.1; -;  
DR HSSP; P01132; 1EGF.  
DR InterPro; IPR000152; -;  
DR InterPro; IPR000436; -;

DR InterPro; IPR000538; -;  
DR InterPro; IPR000561; -;  
DR InterPro; IPR000742; -;  
DR InterPro; IPR001304; -;  
DR InterPro; IPR001438; -;  
DR InterPro; IPR001881; -;  
DR InterPro; IPR003006; -;  
DR Pfam; PF00008; EGF; 2.  
DR Pfam; PF00047; Iq; 1.  
DR Pfam; PF00059; lectin\_c; 1.  
DR Pfam; PF00084; sushi; 1.  
DR Pfam; PF00193; Xlink; 2.  
DR PRINTS; PR00010; EGFBLLOOD.  
DR PRODom; PD000918; -; 2.  
DR PROSITE; PS00010; ASX\_HYDROXYL; UNKNOWN\_1.  
DR PROSITE; PS00615; C-TYPE\_LECTIN\_1; 1.  
DR PROSITE; PS00622; EGF\_1; UNKNOWN\_2.  
DR PROSITE; PS01186; EGF\_2; 1.  
DR PROSITE; PS01187; EGF\_CA; 1.  
DR PROSITE; PS01241; LINK; 2.  
DR SMART; SM00032; CCP; 1.  
DR Calcium-binding; EGF-like domain; Glycoprotein; Repeat; Signal.  
KW SIGNAL  
FT CHAIN 21 656  
SQ SEQUENCE 656 AA; 74793 MW; F8FE153BD10C7AB9 CRC64;  
  
Query Match 88.3%; Score 53; DB 6; Length 656;  
Best Local Similarity 90.0%; Pred. No. 0.16; 1; Indels 0; Gaps 0;  
Matches 9; Conservative 0; Mismatches 1;  
  
Qy 1 CVIGYSGDRC 10  
Db 375 CVPGYSGDRC 384  
I:|||||||  
-:|||||||  
  
RESULT 3  
077611 PRELIMINARY; PRT; 1643 AA.  
AC 077611;  
DT 01-NOV-1998 (TREMBlrel. 08, Created)  
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE VERSICAN V2 SPLICE-VARIANT PRECURSOR.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98288320; PubMed=9624174;  
RA Schmalfeldt M., Dours-Zimmermann M.T., Winterhalter K.H.,  
RT "Versican V2 is a major extracellular matrix component of the mature bovine brain."  
RL J. Biol. Chem. 273:15758-15764(1998).  
CC -|- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX DOMAIN  
DR EMBL; AF060458; AAC24360.1; -;  
DR HSSP; P01132; 1EGF.  
DR InterPro; IPR000152; -;  
DR InterPro; IPR000436; -;  
DR InterPro; IPR000538; -;  
DR InterPro; IPR000561; -;  
DR InterPro; IPR000742; -;  
DR InterPro; IPR001304; -;  
DR InterPro; IPR001438; -;  
DR InterPro; IPR001881; -;  
DR InterPro; IPR003006; -;  
DR Pfam; PF00008; EGF; 2.

```
DR Pfam; PF00047; ig; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; Xlink; 2.
DR PRINTS; PR00010; EGFBLD.
DR ProDom; PD000918; -; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 2.
DR SMART; SM00032; CCP; 1.
DR Calcium-binding; EGF-like domain; Glycoprotein; Repeat; Signal.
KW SMART; SM00032; CCP; 1.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 1643 VERSICAN V2 SPLICE-VARIANT.
SQ SEQUENCE 1643 AA; 182894 MW; A6F2BFC3A3DEF80A CRC64;

Query Match 88.3%; Score 53; DB 6; Length 1643;
Best Local Similarity 90.0%; Pred. No. 0.42;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGSGDRC 10
Db 1362 CVPYSGDRC 1371

RESULT 4
ID O77610 PRELIMINARY; PRT; 2394 AA.
AC O77610;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE VERSICAN V1 SPLICE-VARIANT PRECURSOR.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98288320; PubMed=9624174;
RA Schmalfeldt M., Dours-Zimmermann M.T., Winterhalter K.H.,
RA Zimmermann D.R.;
RT "Versican V2 is a major extracellular matrix component of the mature
RT bovine brain.";
RL J. Biol. Chem. 273:15758-15764(1998).
CC -!- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC DOMAIN.
DR EMBL; AF060457; AAC24359.1; -.
DR HSSP; P01132; LEGF.
DR InterPro; IPR000152; -.
DR InterPro; IPR000436; -.
DR InterPro; IPR000538; -.
DR InterPro; IPR000561; -.
DR InterPro; IPR000742; -.
DR InterPro; IPR001304; -.
DR InterPro; IPR001438; -.
DR InterPro; IPR001881; -.
DR InterPro; IPR003006; -.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00047; ig; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; Xlink; 2.
DR PRINTS; PR00010; EGFBLD.
DR ProDom; PD000918; -; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 2.
DR SMART; SM00032; CCP; 1.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 1643 VERSICAN V0 SPLICE-VARIANT.
SQ SEQUENCE 1643 AA; 182894 MW; A6F2BFC3A3DEF80A CRC64;
```

```
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 2.
DR SMART; SM00032; CCP; 1.
DR Calcium-binding; EGF-like domain; Glycoprotein; Repeat; Signal.
KW SMART; SM00032; CCP; 1.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 2394 VERSICAN V1 SPLICE-VARIANT.
SQ SEQUENCE 2394 AA; 261886 MW; B82A3E10FC5BD990 CRC64;

Query Match 88.3%; Score 53; DB 6; Length 2394;
Best Local Similarity 90.0%; Pred. No. 0.63;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGSGDRC 10
Db 2113 CVPYSGDRC 2122

RESULT 5
ID O77609 PRELIMINARY; PRT; 3381 AA.
AC O77609;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE VERSICAN V0 SPLICE-VARIANT PRECURSOR.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98288320; PubMed=9624174;
RA Schmalfeldt M., Dours-Zimmermann M.T., Winterhalter K.H.,
RA Zimmermann D.R.;
RT "Versican V2 is a major extracellular matrix component of the mature
RT bovine brain.";
RL J. Biol. Chem. 273:15758-15764(1998).
CC -!- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
CC DOMAIN.
DR EMBL; AF060456; AAC24358.1; -.
DR HSSP; P01132; LEGF.
DR InterPro; IPR000152; -.
DR InterPro; IPR000436; -.
DR InterPro; IPR000538; -.
DR InterPro; IPR000561; -.
DR InterPro; IPR000742; -.
DR InterPro; IPR001304; -.
DR InterPro; IPR001438; -.
DR InterPro; IPR001881; -.
DR InterPro; IPR003006; -.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00047; ig; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; Xlink; 2.
DR PRINTS; PR00010; EGFBLD.
DR ProDom; PD000918; -; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 2.
DR SMART; SM00032; CCP; 1.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 3381 VERSICAN V0 SPLICE-VARIANT.
SQ SEQUENCE 3381 AA; 369987 MW; F09716FA778D459 CRC64;
```

Query Match 88.3%; Score 53; DB 6; Length 3381;  
 Best Local Similarity 90.0%; Pred. No. 0.92;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10  
 DB 3100 CVPGYSGDRC 3109

RESULT 6  
 Q21756 PRELIMINARY; PRT; 372 AA.  
 ID Q21756  
 AC Q21756  
 DT 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
 DE HYPOTHETICAL 39.1 KDA PROTEIN.  
 GN R05G66.9.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M., Coulson A.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,  
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans";  
 RL Nature 368:32-38(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX Murray J., Le T.T.;  
 RA Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX Waterston R.;  
 RA Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U58746; AAB00626.1; -;  
 DR InterPro: IPR000561; -;  
 DR Pfam: PF00008; EGF; 6;  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_7.  
 DR PROSITE: PS01186; EGF\_2; 6.  
 DR SMART: SM00181; EGF; 1.  
 KW EGF-like domain; Glycoprotein; Hypothetical protein.  
 FT DOMAIN 90  
 RN [1]  
 RP SEQUENCE 372 AA; 3905 MW; DB36AB251EB6884 CRC64;

Query Match 83.3%; Score 50; DB 5; Length 372;  
 Best Local Similarity 70.0%; Pred. No. 0.3;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10  
 DB 233 CVIGYSGDRC 242

RESULT 7  
 Q28867

ID Q28867 PRELIMINARY; PRT; 89 AA.  
 AC Q28867  
 DT 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
 DE EPIDERMAL GROWTH FACTOR (FRAGMENT).  
 OS Equus caballus (Horse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
 OX NCBI\_TaxID=9796;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95000251; PubMed=7916972;  
 RA Stewart F., Power C.A., Lennard S.N., Allen W.R., Amet L.,  
 RA Edwards R.M.;  
 RT "Identification of the horse epidermal growth factor (EGF) coding  
 RT sequence and its use in monitoring EGF gene expression in the  
 RL endometrium of the pregnant mare";  
 RL J. Mol. Endocrinol. 12:341-350(1994).  
 DR EMBL: S73527; AAB32226.1; -;  
 DR HSSP: P01132; IEGF.  
 DR InterPro: IPR000152; -;  
 DR InterPro: IPR000561; -;  
 DR InterPro: IPR001336; -;  
 DR Pfam: PF00008; EGF; 1.  
 DR PRINTS: PR00009; EGF\_TGF.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; UNKNOWN\_1.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_1.  
 DR PROSITE: PS01186; EGF\_2; 2.  
 DR SMART: SM00181; EGF; 1.  
 KW EGF-like domain; Glycoprotein.  
 FT NON\_TER 1  
 RP SEQUENCE 89 AA; 9879 MW; A7856F5E870B4A4B CRC64;

Query Match 81.7%; Score 49; DB 6; Length 89;  
 Best Local Similarity 70.0%; Pred. No. 0.099;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10  
 DB 74 CVVGYGERC 83

RESULT 8  
 Q61240 PRELIMINARY; PRT; 2352 AA.  
 ID Q61240  
 AC Q61240  
 DT 01-AUG-1998 (TReMBLrel. 07, Created)  
 DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
 DE HRNOTCH PROTEIN.  
 GN HRNOTCH  
 OS Halocynthia roretzi (Sea squirt).  
 OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;  
 OC Stolidobranchia; Pyuridae; Halocynthia.  
 OX NCBI\_TaxID=7729;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Hori S., Saich T., Matsumoto M., Makabe K.W., Nishida H.;  
 RL Dev. Genes Evol. 207:371-380(1997).  
 DR EMBL: AB001327; BAA25571.1; -;  
 DR HSSP: P00740; 1EDM.  
 DR InterPro: IPR000152; -;  
 DR InterPro: IPR000561; -;  
 DR InterPro: IPR000742; -;  
 DR InterPro: IPR000800; -;  
 DR InterPro: IPR001438; -;  
 DR InterPro: IPR001881; -;  
 DR InterPro: IPR002110; -;  
 DR Pfam: PF00008; EGF; 32.  
 DR Pfam: PF00023; ank; 6.  
 DR Pfam: PF00066; notch; 3.

```

DR PRINTS: PR00010; EGFBL00D.
DR PROSITE; PS00088; ANK_REPEAT; 5.
DR PROSITE; PS00297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 18.
DR PROSITE; PS00022; EGF_1; UNKNOWN_28.
DR PROSITE; PS01186; EGF_2; 22.
DR PROSITE; PS01187; EGF_CA; 18.
DR SMART; SM00248; ANK; 1.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
SQ SEQUENCE 2352 AA; 252622 MW; 13DB1C056BBD08D CRC64;

Query Match      80.0%; Score 48; DB 5; Length 2352;
Best Local Similarity 80.0%; Pred. No. 5.1;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10
   || ||||| ||
DB 367 CVAGYSGPRC 376

RESULT 9
ID 035452 PRELIMINARY; PRT; 4006 AA.
AC 035452;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE TENASCIN X.
GN TNX.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowen L., Mahairas G., Qin S., Ahearn M.E., Dankers C., Lasky S.,
RA Loretz C., Schmidt S., Tipton S., Traicoff R., Zackrone K., Hood L.;
RA Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF030001; AAB82015.1; -.
DR HSP; AF02671; IFLD.
DR InterPro; IPR000561; -.
DR InterPro; IPR001777; -.
DR InterPro; IPR002181; -.
DR Pfam; PF00008; EGF; 16.
DR Pfam; PF00041; fn3; 30.
DR Pfam; PF00147; fibrinogen_C; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_19.
DR PROSITE; PS01186; EGF_2; 19.
DR PROSITE; PS00514; FIBRIN_AG_C_DOMAIN; 1.
DR SMART; SM00001; EGF_like; 1.
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 4006 AA; 435476 MW; 553FBB873498A4FC CRC64;

Query Match      80.0%; Score 48; DB 11; Length 4006;
Best Local Similarity 70.0%; Pred. No. 9.1;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10
   | : ||||| |
DB 532 CAVGYSGDC 541

RESULT 10
ID 054796 PRELIMINARY; PRT; 4114 AA.
AC 054796;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE TENASCIN-X.
GN TNX.

```

```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FL BETWEEN C57BL6, AND CBA;
RA Ikuta T., Sogawa N., Ariga H., Ikemura T., Sakakura T., Chiquet-Ehrismann R.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FL BETWEEN C57BL6, AND CBA;
RA MEDLINE=94216385; Pubmed=7512972;
RA Matsumoto K., Saga Y., Ikemura T., Sakakura T., Chiquet-Ehrismann R.;
RT "The distribution of tenascin-X is distinct and often reciprocal to
RT that of tenascin-C.";
RL J. Cell Biol. 125:483-493(1994).
DR EMBL; AB010266; BAA24436.1; -.
DR HSP; P02671; IFLD.
DR InterPro; IPR000561; -.
DR InterPro; IPR001777; -.
DR InterPro; IPR002181; -.
DR Pfam; PF00008; EGF; 16.
DR Pfam; PF00041; fn3; 31.
DR Pfam; PF00147; fibrinogen_C; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_19.
DR PROSITE; PS01186; EGF_2; 17.
DR PROSITE; PS00514; FIBRIN_AG_C_DOMAIN; 1.
DR SMART; SM00060; FN3; 1.
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 4114 AA; 447273 MW; 45FE7AD5145881A1 CRC64;

Query Match      80.0%; Score 48; DB 11; Length 4114;
Best Local Similarity 70.0%; Pred. No. 9.3;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10
   | : ||||| |
DB 532 CAVGYSGDC 541

RESULT 11
ID 09WVB5 PRELIMINARY; PRT; 1531 AA.
AC 09WVB5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE SLIT1.
GN SLIT1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS WEBSTER/ICR;
RA Yuan W., Zhou L., Chen J.-H., Wu J.Y., Rao Y., Ornitz D.M.;
RT "The mouse SLIT family: Secreted ligands for Robo expressed in
RT patterns that suggest a role in morphogenesis and axon guidance.";
RL Dev. Biol. 0:0-0(1999).
DR EMBL; AF144627; AAD44758.1; -.
DR HSP; P00743; ICCF.
DR MGD; MGI:1315203; Slit1.
DR InterPro; IPR000152; -.
DR InterPro; IPR000359; -.
DR InterPro; IPR00372; -.
DR InterPro; IPR000483; -.
DR InterPro; IPR000561; -.
DR InterPro; IPR000742; -.
DR InterPro; IPR001438; -.
DR InterPro; IPR001611; -.

```

```

DR InterPro: IPR001791; -
DR InterPro: IPR001881; -
DR Pfam: PF00008; EGF; 9
DR Pfam: PF00054; laminin_G; 1.
DR Pfam: PF00360; LRR; 19.
DR Pfam: PF01462; LRRNT; 4.
DR Pfam: PF01463; LRRCT; 4.
DR PRINTS: PR00010; EGFBLD.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_2.
DR PROSITE: PS01185; CTCK_1; UNKNOWN_1.
DR PROSITE: PS01225; CTCK_2; 1.
DR PROSITE: PS00022; EGF_1; UNKNOWN_9.
DR PROSITE: PS01186; EGF_2; 8.
DR PROSITE: PS01187; EGF_CA; 2.
DR SMART: SM00013; LRRNT; 1.
KW Calcium-binding; EGF-like domain; Glycoprotein; Repeat.
SQ SEQUENCE 1531 AA; 167545 MW; F7D09AA6693A4F30 CRC64;

Query Match 78.3%; Score 47; DB 11; Length 1531;
Best Local Similarity 60.0%; Pred. No. 4.9;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10
   ||:|||||
Db 1071 CMLGYTGDC 1080

RESULT 12
Q9ROM0 PRELIMINARY; PRT; 2920 AA.
AC Q9ROM0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE FLAMINGO 1.
GN FLAMINGO 1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=1090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99418630; PubMed=10490098;
RA Usui T., Shima Y., Shimada Y., Hirano S., Burgess R.W., Schwarz T.L.,
RA Takeichi M., Uemura T.;
RT "Flamingo, a seven-pass transmembrane cadherin, regulates planar cell
RT polarity under the control of frizzled."
RL Cell 98:585-595(1999).
CC -1- SIMILARITY: TO THE CADHERIN FAMILY.
DR EMBL: AB028499; BAA84070.1; -
DR HSP: P00740; 1IXA.
DR InterPro: IPR000152; -
DR InterPro: IPR000203; -
DR InterPro: IPR000561; -
DR InterPro: IPR000742; -
DR InterPro: IPR000832; -
DR InterPro: IPR001791; -
DR InterPro: IPR001879; -
DR InterPro: IPR002049; -
DR InterPro: IPR002126; -
DR Pfam: PF000002; 7tm_2; 1.
DR Pfam: PF00008; EGF; 5.
DR Pfam: PF00028; cadherin; 9.
DR Pfam: PF00054; laminin_G; 3.
DR Pfam: PF01825; GPS; 1.
DR PRINTS: PR00205; CADHERIN.
DR PRINTS: PR00011; EGFAMININ.
DR PRINTS: PR00249; GPCRSECRETIN.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS00232; CADHERIN; 6.
DR PROSITE: PS00022; EGF_1; UNKNOWN_6.
DR PROSITE: PS01186; EGF_2; 2.

```

```

DR PROSITE: PS01248; LAMININ_TYPE_EGF; UNKNOWN_1.
DR SMART: SM00112; CA; 1.
KW Calcium-binding; Cell adhesion; EGF-like domain; Glycoprotein.
SQ SEQUENCE 2920 AA; 317649 MW; 2919558DF467114F CRC64;

Query Match 78.3%; Score 47; DB 11; Length 2920;
Best Local Similarity 70.0%; Pred. No. 9.9;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10
   ||:|||||
Db 1820 CVLGYGDC 1829

RESULT 13
O29845 PRELIMINARY; PRT; 57 AA.
AC O29845;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE HYPOTHETICAL 6.9 KDA PROTEIN.
GN AF0402.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Kechum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkeess E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:384-370(1997).
DR EMBL: AE001076; AAB90836.1; -
DR TIGR: AF0402; -
KW Hypothetical protein.
SQ SEQUENCE 57 AA; 6885 MW; 4C71BDEB722568B8 CRC64;

Query Match 76.7%; Score 46; DB 1; Length 57;
Best Local Similarity 60.0%; Pred. No. 0.22;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 CVIGYSGDRC 10
   ||:|||||
Db 23 CIMGTGRC 32

RESULT 14
Q9UN93 PRELIMINARY; PRT; 106 AA.
AC Q9UN93;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE MUCIN 3 (FRAGMENT).
GN MUC3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE;
RX MEDLINE=99335363; PubMed=10405327;
RA Williams S.J., Munster D.J., Quin R.J., Gotley D.C., McGuckin M.A.;
RT "The MUC3 gene encodes a transmembrane mucin and is alternatively
   spliced.";
RL Biochem. Biophys. Res. Commun. 261:83-89(1999).
DR EMBL; AF143373; AAD45884.1; -.
DR InterPro; IPR000561; -.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 1.
DR SMART; SM00181; EGF; 1.
DR EGF-like domain; Glycoprotein.
KW NON_TER 1
FT NON_TER 1
SQ SEQUENCE 106 AA; 11340 MW; 5A2E3C24C905E182 CRC64;

Query Match
Best Local Similarity 76.7%; Score 46; DB 4; Length 106;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGYSGDRC 10
Db 14 CLPGFSGDRC 23

RESULT 15
Q9UN94
ID Q9UN94 PRELIMINARY; PRT; 308 AA.
AC Q9UN94;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE MUCIN 3 (FRAGMENT).
EN MUC3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=NORMAL COLONIC MUCOSA;
RX MEDLINE=99335363; PubMed=10405327;
RA Williams S.J., Munster D.J., Quin R.J., Gotley D.C., McGuckin M.A.;
RT "The MUC3 gene encodes a transmembrane mucin and is alternatively
   spliced.";
RL Biochem. Biophys. Res. Commun. 261:83-89(1999).
DR EMBL; AF143372; AAD45883.1; -.
DR InterPro; IPR000082; -.
DR InterPro; IPR000561; -.
DR Pfam; PF01390; SEA; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 1.
DR SMART; SM00181; EGF; 1.
DR EGF-like domain; Glycoprotein.
KW NON_TER 1
FT NON_TER 1
SQ SEQUENCE 308 AA; 34665 MW; A5782A702D4EAE86 CRC64;

Query Match
Best Local Similarity 76.7%; Score 46; DB 4; Length 308;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CVIGYSGDRC 10
Db 14 CLPGFSGDRC 23
```

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2001, 10:54:54 ; Search time 32.64 Seconds  
(without alignments)  
16.716 Million cell updates/sec

Title: US-09-673-785A-4  
Perfect score: 54  
Sequence: 1 CDPGVIGSR 9

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 50623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_0601.\*

1:	/SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2:	/SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3:	/SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4:	/SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5:	/SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6:	/SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7:	/SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8:	/SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9:	/SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10:	/SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11:	/SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12:	/SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13:	/SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14:	/SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15:	/SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16:	/SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17:	/SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18:	/SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19:	/SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20:	/SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21:	/SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22:	/SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	54	100.0	9	AAP82380	Peptide with lamin
2	54	100.0	9	AAP81941	Nonapeptide having
3	54	100.0	9	AAP91162	Peptide with lamin
4	54	100.0	9	AAR08284	Laminin receptor-b
5	54	100.0	9	AAR44011	Cysteine-contg. pe
6	54	100.0	9	AAR44035	Laminin-derived pe
7	54	100.0	9	AAR70490	Cancer metastasis
8	54	100.0	9	AAR92739	YIGSR-containing s
9	54	100.0	9	AAR88569	YIGSR-containing s
10	54	100.0	9	AAB10446	Peptide sequencing
11	54	100.0	9	AAY80486	Cell adhesion pept

12	54	100.0	9	21	AA78851	Laminin fragment u
13	54	100.0	9	21	AA752144	Peptide agonist of
14	54	100.0	9	22	AA70240	Laminin nona-pepti
15	54	100.0	9	22	AA68607	Peptide #1 used in
16	54	100.0	10	21	AA79874	Laminin derived pe
17	54	100.0	40	11	AA07985	LamB1-40. Synthet
18	54	100.0	40	11	AA07990	Laminin B1-40 gene
19	54	100.0	40	11	AA07991	Laminin B1-40 (Ser
20	54	100.0	135	11	AA07984	CTAP(Leu21)/LamB1-
21	54	100.0	164	13	AA22446	Tumour necrosis fa
22	54	100.0	164	13	AA22554	Tumour necrosis fa
23	54	100.0	164	14	AA43026	Human TNF (variant
24	54	100.0	466	11	AA07447	Human laminin B1 c
25	54	100.0	1725	21	AA19800	Mouse laminin 2 ma
26	54	100.0	1725	21	AA19800	Mouse laminin 2 ma
27	54	100.0	1764	10	AA191672	Primary amino acid
28	54	100.0	1765	21	AA19798	Human laminin 2 ma
29	54	100.0	1765	21	AA19798	Human laminin 2 ma
30	54	100.0	1776	19	AA50894	Mouse laminin B1 c
31	54	100.0	1785	20	AA15461	Human laminin beta
32	54	100.0	1786	19	AA50893	Human laminin B1 c
33	54	100.0	1786	21	AA19797	Human laminin 2 be
34	54	100.0	1786	21	AA19797	Human laminin 2 be
35	54	100.0	1786	21	AA19797	Human laminin 2 be
36	54	100.0	1786	21	AA19797	Human laminin 2 be
37	54	100.0	1786	21	AA19797	Human laminin 2 be
38	54	100.0	1786	21	AA19797	Human laminin 2 be
39	50	92.6	49	22	AA72767	Collagen like prot
40	50	92.6	49	22	AA72767	Collagen like prot
41	50	92.6	49	22	AA72767	Collagen like prot
42	50	92.6	69	17	AA95147	Repetitive protein
43	50	92.6	69	22	AA64042	CLP-L1 functional
44	50	92.6	69	22	AA64042	CLP-L1 functional
45	50	92.6	72	17	AA95110	Repetitive protein
			72	22	AA72730	Repetitive protein

#### ALIGNMENTS

RESULT 1

AA82380

ID AAP82380 standard; protein; 9 AA.

XX AC AAP82380;

XX DT 13-NOV-1990 (first entry)

XX DE Peptide with laminin-like activity.

XX KW Laminin; metastasis; wound healing; chemotaxis.

XX OS synthetic.

XX PN EP278781-A.

XX PD 17-AUG-1988.

XX PF 12-FEB-1988; 88EP-0301198.

XX PR 01-OCT-1987; 87US-0102991.

XX PR 12-FEB-1987; 87US-0013919.

XX PR 20-JUL-1988; 88US-0221982.

XX PR 16-NOV-1988; 88US-0272165.

XX (USDC ) US SEC OF COMMERCE.

XX PI Martin GR, Sasaki M, Yamada Y, Kleinman HK, Robey F, Iwamoto Y;

XX PI Graf JO;

XX DR WPI; 1988-229607/33.

XX PT Penta- to nona-peptide(s) having laminin-like activity - having

XX PT an amino acid sequence corresponding to the active domain on the

PT B1 chain of laminin.  
 XX  
 PS Claim 1; Page 11; 27pp; English.  
 XX  
 CC This is an example of a peptide with laminin-like activity e.g.  
 CC ability to promote cell attachment, cell migration and receptor  
 CC elution. It is useful in wound healing, drug targeting and  
 CC inhibiting metastasis in cancer. See also AAP82379 and AAP82381-82.  
 XX  
 XX Sequence 9 AA;  
 SQ

Query Match 100.0%; Score 54; DB 9; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9  
 |||||  
 Db 1 cdpgyigr 9

RESULT 2  
 AAP81941  
 ID AAP81941 standard; peptide; 9 AA.  
 XX  
 AC AAP81941;  
 XX  
 DT 15-OCT-1990 (first entry)  
 XX  
 DE Nonapeptide having laminin-like activity.  
 XX  
 KW laminin-like activity; anti-metastatic agent;  
 KW cell attachment protein.  
 XX  
 OS synthetic.  
 XX  
 PN US7013919-A.  
 XX  
 PD 12-FEB-1988.  
 XX  
 PF 12-FEB-1987; 87US-0013919.  
 XX  
 PR 12-FEB-1987; 87US-0013919.  
 XX  
 PA (USSH ) US DEPT HEALTH AND HUMAN SERVICES.  
 PA (USDC ) US SEC OF COMMERCE.  
 XX  
 PI Yamada Y, Iwamoto Y, Graf J;  
 XX  
 DR WPI; 1988-063855/09.  
 XX  
 PT Peptides having laminin-like activity -  
 PT used as anti-metastatic agents for tumour cells and for  
 PT promoting increased adhesion and cell growth  
 XX  
 PS Claim 1; Page 20; 23pp; English.  
 XX  
 CC Peptides derived by deleting the first 1,2 or 4 amino acid residues  
 CC of this sequence are also claimed. These peptides can all be used  
 CC e.g. as anti-metastatic agents, to target drugs to metastatic  
 CC tumour cells and as cell attachment proteins,  
 CC See also AAP81942-P81944.  
 XX  
 XX Sequence 9 AA;  
 SQ

Query Match 100.0%; Score 54; DB 9; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9  
 |||||  
 Db 1 cdpgyigr 9

RESULT 4  
 AAR08284  
 ID AAR08284 standard; protein; 9 AA.  
 XX  
 AC AAR08284;  
 XX  
 DT 04-MAR-1991 (first entry)  
 XX

RESULT 3  
 AAP91162  
 ID AAP91162 standard; peptide; 9 AA.  
 XX  
 AC AAP91162;  
 XX  
 DT 30-APR-1990 (first entry)  
 XX  
 DE Peptide with laminin activity which blocks angiogenesis.  
 XX  
 KW Laminin activity; angiogenesis blocker; laminin B1 chain;  
 XX US7221982-A.  
 PN  
 XX 07-MAR-1989.  
 PD  
 XX 20-JUL-1988; 88US-0221982.  
 PF  
 XX 20-JUL-1988; 88US-0221982, US-013919.  
 PR  
 XX (USSH ) US DEPT HEALTH & HUMAN.  
 PA  
 XX Yamada Y, Graf JO, Iwamoto Y, Robey FA, Kleinman HK;  
 PI WPI; 1989-150594/20.  
 DR  
 XX Peptide(s) with laminin activity which block angiogenesis  
 PT - alter formation of capillary structures by  
 PT endothelial cells, prevent formation of excess blood  
 PT vessels in tissues etc.  
 XX  
 PS Page 21; ; 24pp; English.  
 XX

It was prepd. as follows. The prim. peptide sequence of one laminin  
 chain was determined by cDNA cloning and the active domain on the B1  
 chain responsible for cell attachment and migration was identified  
 using synthetic peptides. Peptides of 20 amino acids and their corresp.  
 antibodies were prepd. to each of the 7 structural domains. None of  
 these peptides was active although one of the antibodies blocked cell  
 attachment. Smaller synthetic peptides were prepd. to the region around  
 the amino acid sequence specific to this active antibody. Peptide  
 AAP91162 was found to be directly active in cell attachment and cell  
 migration. It blocks angiogenesis, alters the formation of capillary  
 structures by endothelial cells, prevents the formation of excess blood  
 vessels in tissues, and inhibits in vivo tumour cell colonisation of  
 tissues. Other applications are as a carrier of target drugs to  
 metastatic tumour cells, as substrata for cell attachment for in vitro  
 applicns. (such as in cell cultures) and for in vivo applicns., and to  
 promote growth of a specific cell type on a partic. surface prior to  
 grafting. Dosage is 10 microgram - 20 milligram/kg.

Query Match 100.0%; Score 54; DB 10; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9  
 |||||  
 Db 1 cdpgyigr 9

Sequence 9 AA;

```

DE Laminin receptor-binding polypeptide component of Ab-polypeptide
DE conjugate.
XX
XX Integrin-binding polypeptide; laminin-binding polypeptide;
KW Ab-targeting; platelet adhesion; tumour cell adhesion; metastasis.
XX
XX WO9014103-A.
PN
XX
XX 29-NOV-1990.
PD
XX
XX 15-MAY-1990; 90WO-US02746.
XX
XX 17-MAY-1989; 89US-0362617.
PR
XX
XX (SCRI-) SCRIPPS CLINIC & RE.
PA
XX
XX Cheresh DA;
PI
XX
XX WPI; 1990-375777/50.
DR
XX
XX Polypeptide-antibody complex - immuno:reacts with cell surface
PT antigens to inhibit platelet and tumour cell adhesion
PT
XX
XX Disclosure; page 19; 62pp; English.
PS
XX
XX This laminin receptor-binding polypeptide has a sequence contg.
CC YIGSR and is operatively linked to an Ab which targets it to
CC specific sites for inhibiting adhesitory cell attachment (e.g.
CC tumour cells) and platelet adhesion. See also AAR08278-79 and
CC AAR08282-83.
XX
XX SQ Sequence 9 AA;

Query Match 100.0%; Score 54; DB 11; Length 9;
Best Local Similarity 100.0%; Pred. No. 3.4e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9
Db | | | | | | | | |
1 cdpgyigsr 9

RESULT 5
AAR44011
ID AAR44011 standard; peptide; 9 AA.
XX
XX
AC AAR44011;
XX
XX 09-DEC-1993 (first entry)
DT
XX
XX Cysteine-contg. peptide for labelling with a metal ion.
DE
XX
XX integrin; cytoadhesiveness; mast cell; in vivo tumour localisation;
KW imaging; cell attachment; technetium label;
KW medically useful metal ion binding domain.
XX
XX Synthetic.
OS
XX
XX Key Location/Qualifiers
FH 5..9
FT Domain /note= "biological function domain"
FT
XX
XX WO9312819-A.
PN
XX
XX 08-JUL-1993.
PD
XX
XX 31-DEC-1992; 92WO-US11334.
PF
XX
XX 03-JAN-1992; 92US-0816476.
PR
XX
XX 03-JAN-1992; 92US-0816477.
PR
XX
XX 20-FEB-1992; 92US-0840077.
PR
XX
XX 30-DEC-1992; 92US-0998820.
PR

```

```

PR 30-DEC-1992; 92US-0998910.
XX
XX (RHOM-) RHOMED INC.
PA
XX
XX Rhodes BA, Zamora PO;
PI
XX
XX WPI; 1993-227063/28.
DR
XX
XX Metal labelled peptide(s) contg. binding and medically useful
PT domains - used in diagnosis and therapy of e.g. thrombus, cancer,
PT infection, inflammation, are also opt. combined to antibody
PT
XX
XX Example 3; Page 33; 61pp; English.
PS
XX
XX Peptides AAR44011 and AAR44012 were synthesised to compare the
CC potential binding of 99mTc to histidine and cysteine. The invention
CC includes peptides containing one or more amino acids contg. S, O or
CC N, partic the amino acids Cys, His or penicillamine. The peptides
CC were labelled with 99mTc by addition of sodium pertechnetate following
CC treatment with stannous tartrate. The histidine-contg. peptide
CC (AAR44012) bound some but not all the added 99mTc while the
CC cysteine-contg. peptide (AAR44011) bound essentially all the added
CC 99mTc. A poly-Tyrosine control did not bind any label.
XX
XX SQ Sequence 9 AA;

Query Match 100.0%; Score 54; DB 14; Length 9;
Best Local Similarity 100.0%; Pred. No. 3.4e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9
Db | | | | | | | | |
1 cdpgyigsr 9

RESULT 6
AAR44035
ID AAR44035 standard; peptide; 9 AA.
XX
XX
AC AAR44035;
XX
XX 09-DEC-1993 (first entry)
DT
XX
XX Laminin-derived peptide for Tc labelling, contains YIGSR.
DE
XX
XX integrin; cytoadhesiveness; mast cell; anti-metastatic agent;
KW thrombus imaging; platelet adherence; thrombosis; Technetium label;
KW metal ion binding domain; embolism.
XX
XX Synthetic.
OS
XX
XX Key Location/Qualifiers
FH 5..9
FT Domain /label= biological-function_domain
FT /note= "for thrombus imaging"
FT
XX
XX WO9312819-A.
PN
XX
XX 08-JUL-1993.
PD
XX
XX 31-DEC-1992; 92WO-US11334.
PF
XX
XX 03-JAN-1992; 92US-0816476.
PR
XX
XX 03-JAN-1992; 92US-0816477.
PR
XX
XX 20-FEB-1992; 92US-0840077.
PR
XX
XX 30-DEC-1992; 92US-0998820.
PR
XX
XX (RHOM-) RHOMED INC.
PA
XX
XX Rhodes BA, Zamora PO;
PI
XX
XX

```

DR WPI; 1993-227063/28.  
XX  
PT Metal labelled peptide(s) contg. binding and medically useful  
PT domains - used in diagnosis and therapy of e.g. thrombus, cancer,  
PT infection, inflammation, are also opt. combined to antibody  
XX  
XX Example 13; Page 40; 61pp; English.  
XX  
CC This laminin-derived peptide contains the pentapeptide motif YIGSR,  
CC an adhesive sequence from the laminin A-chain which binds to the  
CC 67kd non-integrin platelet receptor. The receptor apparently plays  
CC an important role in the interaction of platelets with the intact  
CC laminin molecule. The sequence is a preferred biological function  
CC domain for peptides of the invention. In addition to a biological  
CC function domain, the peptides contain a metal ion binding domain  
CC (the N-terminal Cys residue in AAR44035) and are labelled with a metal  
CC ion such as Technetium (esp. 99mTc). Peptides comprising YIGSR are  
CC useful for diagnostic imaging of thrombosis and other conditions  
CC characterised by accumulation of platelets.  
XX  
SQ Sequence 9 AA;

Query Match 100.0%; Score 54; DB 14; Length 9;  
Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9  
DB 1 cdpgyigr 9  
|||||

RESULT 7  
AAR70490  
ID AAR70490 standard; peptide; 9 AA.  
AC AAR70490;  
XX  
XX 20-DEC-1995 (first entry)  
DT  
XX  
DE Cancer metastasis inhibitory YIGSR peptide derivative #4.  
XX  
XX Cancer metastasis; adhesive peptide; core sequence; dextran; cancer;  
KW water soluble polysaccharide; metastasis; wound; immunogenicity.  
XX  
XX Synthetic.  
OS  
XX JP07089999-A.  
XX  
XX 04-APR-1995.  
XX  
XX 17-SEP-1993; 93JP-0254779.  
XX  
XX 17-SEP-1993; 93JP-0254779.  
XX  
XX (JAPG ) NIPPON ZEON KK.  
XX  
XX WPI; 1995-167254/22.  
XX  
XX Cancer metastasis inhibitive peptide derivs. - useful for inhibition  
PT of cancer metastasis, healing of wounds and regulation of  
PT immunogenicity.  
XX  
XX Disclosure; Page 3; 6pp; Japanese.  
XX  
XX The peptides AAR70472-90 and AAR82907-24 are peptide derivatives which  
CC inhibit cancer metastasis. They are composed of an adhesive peptide  
CC with a core sequence selected from: RGD (AAR70472-85), YIGSR  
CC (AAR70486-90) or other sequence (AAR82907-24), linked to a water soluble  
CC polysaccharide, preferably a water soluble dextran, at the C-terminus.  
CC The peptides are useful in inhibiting cancer metastasis, healing wounds  
XX and the regulation of immunogenicity.  
XX

SQ Sequence 9 AA;  
Query Match 100.0%; Score 54; DB 17; Length 9;  
Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9  
DB 1 cdpgyigr 9  
|||||

RESULT 8  
AAR92739  
ID AAR92739 standard; peptide; 9 AA.  
XX  
XX AAR92739;  
AC  
XX  
DT 03-SEP-1996 (first entry)  
XX  
DE YIGSR-containing sequence, for controlling cell distribution.  
XX  
XX Control; distribution; bioartificial organ; BAO; cellular attachment;  
KW neurotransmitter; hormone; cytokine; growth factor; enzyme.  
XX  
XX Synthetic.  
OS  
XX  
XX Key Location/Qualifiers  
FH Peptide 5..9  
FT  
FT  
XX WO9602646-A2.  
XX  
XX 01-FEB-1996.  
XX  
XX 20-JUL-1995; 95WO-US09281.  
XX  
XX 09-MAY-1995; 95US-0432698.  
XX 20-JUL-1994; 94US-0279773.  
XX  
XX (CYTO-) CYTOTHERAPEUTICS INC.  
XX  
XX Aebischer P, Cain BM, Doherty EJ, Gentile FT, Hamman JP;  
PI Holland LM, Schinstine M, Shoichet MS, Winn SR;  
XX WPI; 1996-105908/11.  
XX  
XX Controlling distribution of cells in bio-artificial organs - e.g. by  
PT treatment of cells, or growth surfaces, to inhibit proliferation,  
PT promote differentiation or modulate adhesion, for in vivo prodn. of  
PT hormones, neuro-transmitter(s) etc  
XX  
XX Claim 22; Page 70; 84pp; English.  
XX  
XX The sequences given in AAR92739-41 are peptides which were used in the  
CC method of the invention to control the distribution of cells within  
CC a bioartificial organ (BAO). These peptides have been particularly  
CC useful in promoting cellular attachment. These peptides are pref.  
CC bound to the membrane of the BAO which is a biocompatible,  
CC permselective jacket. These peptides act to control the distribution  
CC of the core of living cells included in the BAO after in vivo  
CC implantation. BAO are used therapeutically to produce e.g.  
CC neurotransmitters, hormones, cytokines, growth factors, enzymes, etc.  
XX  
XX  
SQ Sequence 9 AA;  
Query Match 100.0%; Score 54; DB 17; Length 9;  
Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9  
DB 1 cdpgyigr 9  
|||||

```

Db      1 cdpgyigr 9

RESULT  9
AAR88569
ID      AAR88569 standard; peptide; 9 AA.
XX
XX      AAR88569;
AC
XX
XX      04-SEP-1996 (first entry)
DT
XX
XX      YIGSR-containing sequence, for controlling cell distribution.
DE
XX
XX      Laminin-derived peptide; bioartificial; regeneration; nerve;
KW      3-D hydrogel extracellular matrix; proliferation; neurite;
KW      replacement; cartilage; tendon; muscle; bone; skin.
XX
XX      Synthetic.
OS
XX
XX      Key      Location/Qualifiers
FH      Peptide  5..9
FT      /note= "Claimed core peptide, claim 2"
PT
XX
XX      WO9602286-A1.
PN
XX
XX      01-FEB-1996.
PD
XX
XX      20-JUL-1995; 95WO-US09282.
PF
XX
XX      20-JUL-1994; 94US-0280646.
PR
XX
XX      (CYTO-) CYTOTHERAPEUTICS INC.
PA
XX
XX      Aebischer P, Bellamkonda RV, Ranieri JP;
PI
XX
XX      WPI: 1996-105660/11.
DR
XX
XX      Bio-artificial 3-D hydrogel extracellular matrix comprising hydrogel
PT      derivatised with adhesion molecules - useful for promoting in vivo
PT      regeneration of severed nerves, tissue replacement and cell
PT      manipulation
XX
XX      Claim 3; Page 49; 65pp; English.
PS
XX
XX      The sequences given in AAR88569-71 are laminin-derived peptides which
CC      were used in the bioartificial 3-D hydrogel extracellular matrix
CC      of the invention to control the distribution of cells. These peptides
CC      are particularly useful in promoting cellular proliferation in neurites.
CC      These peptides are used to derivatise the hydrogel. The hydrogel is a
CC      polysaccharide and has a pore radius of > 120 nm, pref. 150 nm. The
CC      hydrogel is useful for promoting in vivo regeneration of a severed
CC      nerve. It may have cells suspended in it and may be used to promote
CC      in vivo replacement of cartilage, tendon, muscle, bone or skin.
XX
XX      Sequence 9 AA;

Query Match      100.0%; Score 54; DB 17; Length 9;
Best Local Similarity 100.0%; Pred. No. 3.4e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CDPGYIGSR 9
        |||||
Db      1 cdpgyigr 9

RESULT  10
AAB10446
ID      AAB10446 standard; peptide; 9 AA.
XX
XX      AAB10446;
AC
XX
XX      01-DEC-2000 (first entry)
DT

```

```

XX
DE
XX
KW      Peptide sequencing method synthetic peptide SEQ ID NO: 2.
KW      post-translational modification; laundry; cleansing product; proteomic;
KW      y-ion.
XX
OS      Synthetic.
XX
XX      WO200043792-A2.
PN
XX
XX      27-JUL-2000.
PD
XX
XX      12-JAN-2000; 2000WO-US00790.
PF
XX
XX      20-JAN-1999; 99US-0116502.
PR
XX
XX      29-SEP-1999; 99US-0156677.
XX
XX      (PROC ) PROCTER & GAMBLE CO.
PA
XX
XX      Keough TW, Youngquist RS;
PI
XX
XX      WPI: 2000-543265/49.
DR
XX
XX      Determining amino acid sequence of polypeptide by derivatizing the
PT      N-terminus of the polypeptide with acidic moieties, analyzing
PT      derivatized products using mass spectrometric technique and
PT      interpreting the fragmentation pattern -
XX
XX      Example 4; Page 29; 30pp; English.
PS
XX
XX      This invention describes a novel method for determining the amino acid
CC      sequence of a polypeptide comprising derivatizing the N-terminus of the
CC      polypeptide or polypeptides with one or more acidic moieties with pKa of
CC      less than 2 when coupled with the polypeptide or polypeptides, analyzing
CC      the derivatized products using a mass spectrometric technique to provide
CC      a fragmentation pattern free of a- and b-ions and interpreting the
CC      fragmentation pattern. The method is used for sequencing wild-type or
CC      variant polypeptides. Applications include biological studies,
CC      identification of post-translational modifications in proteins,
CC      identification of amino acid modifications in variant proteins used in
CC      e.g. commercial laundry and cleansing products, designing oligonucleotide
CC      probes for gene cloning, rapid characterization of products formed in
CC      directed evolution studies, combinatorial chemistry and peptide libraries
CC      and proteomics. Derivatization of the polypeptides with acid groups gives
CC      almost exclusive y-ion fragmentation and very little a-ion and b-ion
CC      'noise', providing mass spectra which are more easily interpreted. The
CC      method is simple, efficient and widely applicable to both wild-type and
CC      variant polypeptides. This sequence represents a synthetic peptide used
CC      to illustrate the method of the invention.
XX
XX      Sequence 9 AA;

Query Match      100.0%; Score 54; DB 21; Length 9;
Best Local Similarity 100.0%; Pred. No. 3.4e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 CDPGYIGSR 9
        |||||
Db      1 cdpgyigr 9

RESULT  11
AAY80486
ID      AAY80486 standard; peptide; 9 AA.
XX
XX      AAY80486;
AC
XX
XX      06-JUN-2000 (first entry)
DT
XX
XX      Cell adhesion peptide #21.
DE
XX

```

KW Bone regenerative; osteopathic; osseous tissue; reconstitution;  
 KW scaffold matrix; bone formation promoter; bone resorption inhibitor;  
 KW cell adhesion; osteoblast; osteoclast; bone defect; fracture.

XX Synthetic.

OS WO200004941-A1.

PN 03-FEB-2000.

XX 22-JUL-1999; 99WO-US16800.

XX 24-JUL-1998; 98US-0122348.

XX (PHAR-) PHARMACAL BIOTECHNOLOGIES INC.

XX Budny JA;

XX WPI; 2000-195084/17.

XX System for reconstructing osseous tissue, useful e.g. for treating  
 PT fractures, comprises scaffold containing promoter of bone formation and  
 PT inhibitor of bone resorption.

XX Claim 14; Page 32; 44pp; English.

XX The invention relates to a novel system for reconstitution of osseous  
 CC tissue comprising a scaffold carrying a compound (I) that promotes  
 CC bone formation and a component that decreases bone resorption (II).  
 CC (I) induces migration and adhesion of osteoblasts and osteoclasts and  
 CC (II) inhibits proteolysis (specifically by plasmin) of extracellular  
 CC matrix. (I) is preferably selected from: selectin or selectin binding  
 CC fragments, proteins and peptides that facilitate cell adhesion,  
 CC plasminogen activator inhibitors, protease inhibitors and  
 CC metalloprotease inhibitors. The peptides AAR80466-Y80492 are claimed  
 CC examples of cell adhesion peptides used in the system of the invention.  
 CC The system is used to replace, remodel or correct bone defects, e.g.  
 CC fractures, fissures or bone mass loss. Incorporation of (I) into the  
 CC scaffold results in rapid seeding by osteoblasts and the development of  
 CC an organic matrix, i.e. the preformed scaffold replaces the  
 CC rate-determining step of extracellular matrix formation. The scaffold can  
 CC be designed to have a predetermined resorption/degradation rate, and may  
 CC include regulatory compounds for specific cell types.

XX Sequence 9 AA;

Query Match 100.0%; Score 54; DB 21; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9

Db | | | | | | | | | |

Db 1 cdpgyigr 9

RESULT 12

AY78851

ID AAY78851 standard; Peptide; 9 AA.

XX AAY78851;

DT 19-MAY-2000 (first entry)

XX Laminin fragment used as a fusion protein acceptor.

DE Laminin; peptide production; transgenic animal; production in milk.

XX Unidentified.

XX Key Location/Qualifiers

FT Modified-site 9 /note= "C-terminal amide"

FT

XX WO200000625-A1.  
 PN 06-JAN-2000.  
 PD 16-JUN-1999; 99WO-GB01907.  
 PF 26-JUN-1998; 98GB-0013912.  
 PR 28-AUG-1998; 98US-0098281.

XX (PPLT-) PPL THERAPEUTICS SCOTLAND LTD.

XX Cottingham IR, McKee CM, Millar AR;

XX WPI; 2000-170922/15.

XX Novel methods for production of peptides with authentic amino-termini,  
 PT and peptide-acceptor conjugates.

XX Example 3; Page 25; 43pp; English.

XX This sequence represents an laminin peptide which is used as an acceptor  
 CC for the fusion proteins of the invention. The invention relates to the  
 CC production of a peptide with an authentic amino terminal comprising  
 CC expressing the peptide as part of a fusion protein, wherein the peptide  
 CC incorporates a sequence extension at its N-terminus. The fusion protein  
 CC of the invention can be used in the production of transgenic animals  
 CC which produce the desired peptide in their milk. The methods of the  
 CC invention are used to make peptides with authentic amino terminal amino  
 CC acids, and peptide-acceptor conjugates.

XX Sequence 9 AA;

Query Match 100.0%; Score 54; DB 21; Length 9;

Best Local Similarity 100.0%; Pred. No. 3.4e+05;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9

Db | | | | | | | | | |

Db 1 cdpgyigr 9

RESULT 13

AY52144

ID AAY52144 standard; peptide; 9 AA.

XX AAY52144;

XX 28-JAN-2000 (first entry)

DT Peptide agonist of laminin receptor.

DE Epidermal growth factor; EGF; laminin receptor; angiogenesis;

XX medicament; wound healing; retinopathy of immaturity; metastatic cancer;  
 KW candida infection; leishmaniasis; trichomonas vaginalis.  
 KW Mus SP.

OS Key

FT Location/Qualifiers

FT Modified-site 9 /note= "C-terminal amide"

XX WO9954356-A1.

XX 28-OCT-1999.

XX 21-APR-1999; 99WO-GB01211.

XX 22-APR-1998; 98GB-0008407.

XX (UYBE-) UNIV QUEENS BELFAST.

XX

PI Nelson J, Walker B, McFerran N, Harriott P;

XX WPI; 2000-013229/01.

XX New peptide derived from murine epidermal growth factor (MEGF)

XX Example 1; Page 5; 35pp; English.

XX This is a peptide which can be used as an agonist of the mouse laminin  
 CC receptor. The peptide is used in the invention which relates to a  
 CC peptide (AAV52143) derived from mouse epidermal growth factor (EGF)  
 CC residues 33-42. This peptide is used in the invention to prepare a  
 CC composition to target laminin receptors. EGF derived peptides inhibit  
 CC blood vessel formation through their antagonism of the high affinity 67kD  
 CC laminin receptor found on endothelial cells. The peptide is modified from  
 CC the natural sequence to prevent protease attack. The peptide is used in  
 CC the preparation of a medicament for binding to laminin receptors as an  
 CC (ant)agonist. The medicament is also useful for healing endothelial cell  
 CC wounds and treating angiogenic diseases, especially retinopathy of  
 CC immaturity. Other diseases treated include metastatic cancer, Candida  
 CC spp. infection, and parasitic infestations like leishmania and  
 CC trichomonas vaginalis. The peptide are anti-angiogenic in human models.  
 CC The peptides also inhibit both laminin- and EGF-stimulated angiogenesis,  
 CC and prevent tumour cell attachment to basement membranes.

XX Sequence 9 AA;

Query Match 100.0%; Score 54; DB 21; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9  
 |||||

Db 1 cdpgyigr 9

RESULT 14

AA070240  
 ID AAB70240 standard; peptide; 9 AA.

AC AAB70240;

XX 04-MAY-2001 (first entry)

XX Laminin nona-pptide.

XX Chimeric; disulphide; multimeric.

XX Synthetic.

XX EPI074563-A1.

XX 07-FEB-2001.

XX 02-AUG-1999; 99EP-0115022.

XX 02-AUG-1999; 99EP-0115022.

XX (HOFF ) HOFFMANN LA ROCHE & CO AG F.

XX WPI; 2001-184353/19.

XX New chimeric polypeptide, useful as a multimeric pharmaceutical agent,  
 PT comprises polypeptide chains linked by 1-3 cysteine-based disulfide  
 PT bridges and are linked to a biologically active compound -

XX Example 1; Page 7; 26pp; English.

XX The present invention relates to a chimeric polypeptide consisting  
 CC of two polypeptide chains chemically linked through 1-3  
 CC cysteine-based disulphide bridges and linked at its C- or N-  
 CC terminus to a biologically active compound. The invention is

CC useful as a multimeric pharmaceutical agent, by coupling  
 CC biologically active compounds to protein or other agents.

SQ Sequence 9 AA;

Query Match 100.0%; Score 54; DB 22; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9  
 |||||

Db 1 cdpgyigr 9

RESULT 15

AA068607  
 ID AAB68607 standard; peptide; 9 AA.

XX AAB68607;

XX 25-APR-2001 (first entry)

XX Peptide #1 used in a liposome composition.

XX Liposome composition; HIV infection; septic shock; toxic shock;  
 KW colonic inflammation; leukaemic cell proliferation.

XX Unidentified.

XX US6180134-B1.

XX 30-JAN-2001.

XX 07-JUN-1995; 95US-0480332.

XX 23-MAR-1993; 93US-0035443.

XX 29-SEP-1994; 94US-0316436.

XX (SEQU-) SEQUUS PHARM INC.

XX Zalipsky S, Woodie MC, Martin FJ, Barenholz Y;

XX WPI; 2001-201897/20.

XX Liposome composition for use in treating septic shock comprises  
 PT liposomes having an outer surface layer of polyethylene glycol chains,  
 PT and a polypeptide or polysaccharide effector molecule -

XX Disclosure; Fig 13; 32pp; English.

XX The present invention relates to a liposome composition comprising  
 CC liposomes having an outer surface layer of polyethylene glycol chains,  
 CC each having a free distal end. A polypeptide or polysaccharide effector  
 CC molecule is covalently attached to a portion of the distal ends. The  
 CC effector interferes with specific binding of pathogen or cell in a  
 CC bloodstream to a target cell or cell matrix, and is rapidly removed by  
 CC renal clearance from the bloodstream when administered in free form.  
 CC The liposome composition may be used in treating a condition mediated by  
 CC binding a pathogen or cell in the bloodstream, to a target cell or cell  
 CC matrix. It can be used in treating septic shock, toxic shock, colonic  
 CC inflammation, leukaemic cell proliferation, or HIV infection. The present  
 CC sequence is a peptide which may be used in the composition of the present  
 CC invention.

SQ Sequence 9 AA;

Query Match 100.0%; Score 54; DB 22; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 3.4e+05;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9

Db 1 cdpgyigr 9

Search completed: August 15, 2001, 10:54:55  
Job time: 208 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2001, 10:54:15 ; Search time 22.89 Seconds  
(without alignments)  
29,951 Million cell updates/sec

Title: US-09-673-785A-4

Perfect score: 54

Sequence: 1 CDPGVIGSR 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:\*  
1: piri:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54	100.0	1786	1 MMHUB1	laminin beta-1 cha
2	54	100.0	1786	1 MMHUB1	laminin beta-1 cha
3	43	79.6	178	2 JC1467	betacellulin precu
4	43	79.6	656	2 JC2005	integrin beta-5 ch
5	43	79.6	799	2 A38308	hypothetical prote
6	42	77.8	1827	2 T34288	hypothetical prote
7	41	75.9	372	2 T29359	hypothetical prote
8	41	75.9	1620	2 T27283	hypothetical prote
9	40	74.1	372	2 S23936	L-selectin precurs
10	40	74.1	2531	2 S18188	notch protein homo
11	40	74.1	2531	2 A46019	Notch-1 protein -
12	40	74.1	2555	2 A40043	notch protein homo
13	40	74.1	2703	1 A24420	notch protein - fr
14	40	74.1	3566	1 A40701	tenascin-X precurs
15	40	74.1	4135	2 T42629	tenascin-X - bovin
16	39	72.2	79	2 A82667	single-stranded DN
17	39	72.2	761	2 H85579	probable enzyme yb
18	39	72.2	810	2 T10756	Nel-homolog protei
19	39	72.2	1295	2 A32901	glp1 protein precu
20	39	72.2	1574	2 T13954	MEGF6 protein - ra
21	39	72.2	1964	2 T09059	notch4 - mouse
22	39	72.2	2437	2 S42612	transmembrane prot
23	39	72.2	3461	2 S58870	reelin precursor -
24	39	72.2	4006	2 T09070	probable tenascin
25	38	70.4	227	2 S44836	K02D10.4 protein -
26	38	70.4	252	2 T46247	hypothetical prote
27	38	70.4	372	1 UHHUCN	ciliary neurotroph
28	38	70.4	378	2 B59180	Wnt inhibitory fac
29	38	70.4	387	2 B49175	Notch A protein -

30	38	70.4	740	2 F71369	conserved hypothet
31	38	70.4	861	2 A48825	Notch homolog Motc
32	38	70.4	915	2 T21773	hypothetical prote
33	38	70.4	927	2 T21772	hypothetical prote
34	38	70.4	955	2 A45441	thrombospondin 4 -
35	38	70.4	1025	2 T42626	secreted leucine-r
36	38	70.4	1203	2 A49175	Notch B protein -
37	38	70.4	1216	2 T26104	hypothetical prote
38	38	70.4	1469	2 B36665	slit protein 2 pre
39	38	70.4	1480	2 A36665	hypothetical prote
40	38	70.4	1647	2 T41267	hypothetical prote
41	38	70.4	2120	2 T30243	alpha tectorin - c
42	38	70.4	2321	2 S78549	notch3 protein - h
43	38	70.4	2471	2 A49128	cell-fate determin
44	38	70.4	2524	2 A35844	Notch protein - Af
45	38	70.4	3375	2 T19821	hypothetical prote

ALIGNMENTS

RESULT 1  
MMHUB1  
laminin beta-1 chain precursor - human  
N:Alternate names: laminin chain B1  
C:Species: Homo sapiens (man)  
C:Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 19-Jan-2001  
C:Accession: S13547; A28483; A26994; S23566  
R:Vuolteenaho, R.; Chow, L.T.; Tryggvason, K.  
J. Biol. Chem. 265, 15611-15616, 1990  
A:Title: Structure of the human laminin B1 chain gene.  
A:Reference number: S13547; MUID:90368768  
A:Accession: S13547  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1786 <VUO>  
A:Cross-references: GB:M61951; GB:J02778; NID:gl86911; PIDN:AAA59486.1; PID:gl86913  
R:Piikkarainen, T.; Eddy, R.; Fukushima, Y.; Byers, M.; Shows, T.; Pihlajaniemi, T.; S.  
J. Biol. Chem. 262, 10454-10462, 1987  
A:Title: Human laminin B1 chain. A multidomain protein with gene (LAMB1) locus in the  
A:Reference number: A28483; MUID:87280097  
A:Accession: A28483  
A:Molecule type: mRNA  
A:Residues: 1-1786 <PIK>  
A:Cross-references: GB:M61951; GB:J02778; NID:gl86911; PIDN:AAA59486.1; PID:gl86913  
R:Jaye, M.; Modi, W.S.; Rices, G.A.; Mudd, R.; Chiu, I.M.; O'Brien, S.J.; Drohan, W.N.  
Am. J. Hum. Genet. 41, 603-615, 1987  
A:Title: Isolation of a cDNA clone for the human laminin-B1 chain and its gene locali  
A:Reference number: A26994; MUID:88021029  
A:Accession: A26994  
A:Molecule type: mRNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-1709 <JAY>  
A:Cross-references: EMBL:M20206; NID:gl86914; PIDN:AAA59487.1; PID:gl86915  
R:Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Piikkarainen, T.; Tryggvason, K.  
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic  
A:Title: Genes for the human laminin B1 and B2 chains.  
A:Reference number: S23566  
A:Accession: S23566  
A:Molecule type: DNA  
A:Residues: 1276-1469, 'V', 1471-1695, 'G', 1697-170

F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-1786/Product: laminin beta-1 chain #status predicted <MAT>  
 F:22-270/Domain: VI <DOM6>  
 F:271-548/Domain: V <DOM5>  
 F:335-395/Domain: laminin-type EGF-like homology <LE01>  
 F:398-455/Domain: laminin-type EGF-like homology <LE02>  
 F:458-507/Domain: laminin-type EGF-like homology <LE03>  
 F:463-468/Region: cell adhesion #status predicted  
 F:510-540/Domain: laminin-type EGF-like homology <LE04>  
 F:549-774/Domain: IV <DOM4>  
 F:662-668/Region: cell adhesion #status predicted  
 F:773-818/Domain: laminin-type EGF-like homology <LE06>  
 F:773-1178/Domain: III <DOM3>  
 F:821-864/Domain: laminin-type EGF-like homology <LE07>  
 F:867-914/Domain: laminin-type EGF-like homology <LE08>  
 F:917-973/Domain: laminin-type EGF-like homology <LE09>  
 F:923-927/Region: cell adhesion #status predicted  
 F:950-954/Region: cell adhesion #status predicted  
 F:976-1025/Domain: laminin-type EGF-like homology <LE10>  
 F:1028-1081/Domain: laminin-type EGF-like homology <LE11>  
 F:1084-1129/Domain: laminin-type EGF-like homology <LE12>  
 F:1132-1176/Domain: laminin-type EGF-like homology <LE13>  
 F:1179-1397/Domain: II <DOM2>  
 F:1179-1397/Region: heptad repeats  
 F:1398-1430/Domain: alpha <ALP>  
 F:1431-1786/Domain: I <DOM1>  
 F:30-35/Disulfide bonds: #status predicted  
 F:120-356,519,677,965,1041,1195,1279,1336,1343,1487,1542,1643/Binding site: carbohydrate  
 F:1179,1182,1785/Disulfide bonds: Interchain #status predicted

Query Match 100.0%; Score 54; DB 1; Length 1786;  
 Best Local Similarity 100.0%; Pred. No. 0.19;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9  
 |||||  
 Db 946 CDPGYIGSR 954

RESULT 2

MM581  
 laminin beta-1 chain precursor - mouse  
 N:Alternate names: laminin chain B1  
 C:Species: Mus musculus (house mouse)  
 C:Date: 28-Feb-1986 #sequence, revision 30-Jun-1991 #text\_change 10-Dec-1999  
 C:Accession: A26413; S02679; S05326; S14877; A02871; S02036; S13543  
 R:Sasaki, M.; Kato, S.; Komo, K.; Martin, G.R.; Yamada, Y.  
 Proc. Natl. Acad. Sci. U.S.A. 84, 935-939, 1987  
 A:Title: Sequence of the cDNA encoding the laminin B1 chain reveals a multidomain protein  
 A:Reference number: A26413; MUID:87147212  
 A:Accession: A26413  
 A:Molecule type: mRNA  
 A:Residues: 1-1786 <SAS>  
 A:Cross-references: EMBL:M15525; NID:g198700  
 A:Note: translation in GenBank has additional 48 residues at the amino end  
 R:Fujiwara, S.; Shinkai, H.; Deutzmann, R.; Paulsson, M.; Timpl, R.  
 Biochem. J. 252, 453-461, 1988  
 A:Title: Structure and distribution of N-linked oligosaccharide chains on various domain  
 A:Reference number: S02678; MUID:88326259  
 A:Accession: S02679  
 A:Molecule type: protein  
 A:Residues: 28-42; 932-946 <FUJ>  
 R:Harti, L.; Oberbauer, I.; Deutzmann, R.  
 Eur. J. Biochem. 173, 629-635, 1988  
 A:Title: The N terminus of laminin A chain is homologous to the B chains.  
 A:Reference number: S00624; MUID:88225080  
 A:Accession: S05326  
 A:Molecule type: protein  
 A:Residues: 457-466; 854-868; 932-946 <HAR>  
 R:Mann, K.; Deutzmann, R.; Timpl, R.

Eur. J. Biochem. 178, 71-80, 1988  
 A:Title: Characterization of proteolytic fragments of the laminin-nidogen complex and  
 A:Reference number: S08895; MUID:89078415  
 A:Accession: S14877  
 A:Molecule type: protein  
 A:Residues: 590-620 <MAN>  
 R:Barlow, D.P.; Green, N.M.; Kurkinen, M.; Hogan, B.L.M.  
 EMBO J. 3, 2355-2362, 1984  
 A:Title: Sequencing of laminin B chain cDNAs reveals C-terminal regions of coiled-coil  
 A:Reference number: A02870; MUID:85051302  
 A:Accession: A02871  
 A:Molecule type: mRNA  
 A:Residues: 1292-1530, 'MBMP', 1535-1691, 'C', 1693-1748, 'N', 1750-1786 <BAR>  
 A:Cross-references: EMBL:X05212; NID:g52861; PIDN:CAA28839.1; PID:g809042  
 Eur. J. Biochem. 177, 35-45, 1988  
 A:Title: Structural study of long arm fragments of laminin. Evidence for repetitive C  
 A:Reference number: S01790; MUID:89030693  
 A:Accession: S02036  
 A:Molecule type: protein  
 A:Residues: 1561-1587 <DEU>  
 R:Paulsson, M.; Deutzmann, R.; Timpl, R.; Dalzoppo, D.; Odermatt, E.; Engel, J.  
 EMBO J. 4, 309-316, 1985  
 A:Title: Evidence for coiled-coil alpha-helical regions in the long arm of laminin.  
 A:Reference number: S13543; MUID:85257455  
 A:Accession: S13543  
 A:Molecule type: protein  
 A:Residues: 1700-1748, 'N', 1750-1759 <PAU>  
 C:Genetics:  
 A:Gene: Lamb-1  
 A:Map position: 12  
 C:Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type lamin.  
 C:Function:  
 A:Description: Interact with cells and with other basement membrane proteins to promote  
 C:Superfamily: laminin beta-1 chain; laminin-type EGF-like homology  
 C:Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellular  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-1786/Product: laminin beta-1 chain #status predicted <MAT>  
 F:22-270/Domain: VI <DOM6>  
 F:271-540/Domain: V <DOM5>  
 F:271-332/Domain: laminin-type EGF-like homology <LE01>  
 F:335-395/Domain: laminin-type EGF-like homology <LE02>  
 F:398-455/Domain: laminin-type EGF-like homology <LE03>  
 F:458-507/Domain: laminin-type EGF-like homology <LE04>  
 F:510-540/Domain: laminin-type EGF-like homology <LE05>  
 F:541-772/Domain: IV <DOM4>  
 F:773-1182/Domain: III <DOM3>  
 F:773-818/Domain: laminin-type EGF-like homology <LE06>  
 F:821-864/Domain: laminin-type EGF-like homology <LE07>  
 F:867-914/Domain: laminin-type EGF-like homology <LE08>  
 F:917-973/Domain: laminin-type EGF-like homology <LE09>  
 F:976-1025/Domain: laminin-type EGF-like homology <LE10>  
 F:1028-1081/Domain: laminin-type EGF-like homology <LE11>  
 F:1084-1129/Domain: laminin-type EGF-like homology <LE12>  
 F:1132-1176/Domain: laminin-type EGF-like homology <LE13>  
 F:1183-1397/Domain: II <DOM2>  
 F:1183-1397/Region: heptad repeats  
 F:1398-1430/Domain: alpha <ALP>  
 F:1431-1786/Region: heptad repeats  
 F:1431-1786/Domain: I <DOM1>  
 F:22/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted  
 F:30-35/Disulfide bonds: #status predicted  
 F:120-356,519,677,1041,1195,1279,1336,1343,1487,1533,1542,1643/Binding site: carbohydrate  
 F:1179,1182,1785/Disulfide bonds: Interchain #status predicted

Query Match 100.0%; Score 54; DB 1; Length 1786;  
 Best Local Similarity 100.0%; Pred. No. 0.19;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9  
 |||||  
 Db 946 CDPGYIGSR 954

```

RESULT 3
JC1467
betacellulin precursor - human
C:Species: Homo sapiens (man)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999
C:Accession: JC1467
R:Sasada, R.; Ono, Y.; Taniyama, Y.; Shing, Y.; Folkman, J.; Igarashi, K.
Biochem. Biophys. Res. Commun. 190, 1173-1179, 1993
A:Title: Cloning and expression of cDNA encoding human betacellulin, a new member of the
A:Reference number: JC1467; MUID:93176165
A:Accession: JC1467
A:Molecule type: mRNA
A:Residues: 1-178 <SAS>
A:Cross-references: GB:S55606; NID:g265785; PIDN:AAB25452.1; PID:g265786
C:Superfamily: transforming growth factor alpha precursor; EGF homology
C:Keywords: glycoprotein; transmembrane protein
F:1-31/Domain: signal sequence #status predicted <SIG>
F:32-111/Product: betacellulin #status predicted <MAT>
F:69-104/Domain: EGF homology <EGF>
F:119-138/Domain: transmembrane #status predicted <TM>
F:34/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 79.6%; Score 43; DB 2; Length 178;
Best Local Similarity 77.8%; Pred. No. 1.8;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9
|| |||||:|
Db 95 CDEGYIGAR 103

RESULT 4
JC2005
integrin beta-5 chain - baboon
C:Species: Papio sp. (baboon)
C:Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 21-Jan-2000
C:Accession: JC2005
R:Shoji, M.; Hayzer, D.J.; Kim, T.M.; Runge, M.S.; Hanson, S.R.
Gene 133, 307-308, 1993
A:Title: Human and baboon integrin beta 5 subunit-encoding mRNAs have alternative polyad
A:Reference number: JC2005; MUID:94040831
A:Accession: JC2005
A:Molecule type: mRNA
A:Residues: 1-656 <SHO>
A:Note: the authors translated the codon AGA for residue 454 as Lys, CAA for residue 471
C:Comment: This protein is a predominant subunit for the vitronectin receptor in baboon
C:Superfamily: integrin beta chain; laminin-type EGF-like homology
C:Keywords: cell adhesion
F:320-370/Domain: laminin-type EGF-like homology <LEG>

Query Match 79.6%; Score 43; DB 2; Length 656;
Best Local Similarity 66.7%; Pred. No. 6.4;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9
| |||||:|
Db 346 CSPGYIGR 354

RESULT 5
A38308
integrin beta-5 chain precursor - human
C:Species: Homo sapiens (man)
C:Date: 14-Jun-1991 #sequence_revision 14-Jun-1991 #text_change 20-Aug-1999
C:Accession: A38308; A35775; S12534; S11708
R:McLean, J.W.; Vestal, D.J.; Cheresih, D.A.; Bodary, S.C.
J. Biol. Chem. 265, 17126-17131, 1990
A:Title: cDNA sequence of the human integrin beta-5 subunit.
A:Reference number: A38308; MUID:91009141

```

```

A:Accession: A38308
A:Molecule type: mRNA
A:Residues: 1-799 <MCL>
A:Cross-references: GB:J05633; NID:g186504; PIDN:AAA59183.1; PID:g186505
A:Note: parts of this sequence, including the amino end of the mature protein, were c
R:Suzuki, S.; Huang, Z.S.; Tanihara, H.
Proc. Natl. Acad. Sci. U.S.A. 87, 5354-5358, 1990
A:Title: Cloning of an integrin beta subunit exhibiting high homology with integrin b
A:Reference number: A35775; MUID:90319111
A:Accession: A35775
A:Molecule type: mRNA
A:Residues: 1-192, 'A', 194-644, 'L', 646-789, 793-799 <SUZ>
A:Cross-references: GB:M35011; NID:g184524; PIDN:AAAS2707.1; PID:g306894
R:Ramawamy, H.; Hemler, M.E.
EMBO J. 9, 1561-1568, 1990
A:Title: Cloning, primary structure and properties of a novel human integrin beta sub
A:Reference number: S12534; MUID:90228356
A:Accession: S12534
A:Molecule type: mRNA
A:Residues: 1-644, 'L', 646-799 <RAM>
A:Cross-references: EMBL:X53002; NID:g33952; PIDN:CAA37188.1; PID:g33953
C:Genetics:
A:Gene: GDB:ITGB5
A:Cross-references: GDB:128005; OMIM:147561
A:Map position: 17q11-17qter
C:Superfamily: integrin beta chain; laminin-type EGF-like homology
C:Keywords: cell adhesion; cytoskeleton; duplication; extracellular matrix; glycoprot
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-799/Product: integrin beta-5 chain #status experimental <MAT>
F:25-719/Domain: extracellular #status predicted <EXT>
F:463-513/Domain: laminin-type EGF-like homology <LEG>
F:720-742/Domain: transmembrane #status predicted <TM>
F:743-799/Domain: intracellular #status predicted <INT>
F:347,460,477,505,552,586,654,705/Binding site: carbohydrate (Asn) (covalent) #status

Query Match 79.6%; Score 43; DB 2; Length 799;
Best Local Similarity 66.7%; Pred. No. 7.8;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9
| |||||:|
Db 489 CSPGYIGTR 497

RESULT 6
T34288
hypothetical protein F47C12.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34288
R:Murray, J.; Wohldmann, P.; Sansone, J.
submitted to the EMBL Data Library, June 1996
A:Description: The sequence of C. elegans cosmid F47C12.
A:Reference number: Z21499
A:Accession: T34288
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1827 <MUR>
A:Cross-references: EMBL:U61946; PIDN:AAC24388.1; GSPDB:GN00022; CESP:F47C12.1
A:Experimental source: strain Bristol N2; clone F47C12
C:Genetics:
A:Gene: CESP:F47C12.1
A:Map position: 4
A:Introns: 66/2; 221/1; 282/1; 404/1; 500/1; 649/3; 750/1; 997/2; 1152/3; 1279/1; 134

Query Match 77.8%; Score 42; DB 2; Length 1827;
Best Local Similarity 66.7%; Pred. No. 27;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9
| |||||:|

```

Db 1695 CKPGYVGR 1703

RESULT 7

T29359  
hypothetical protein R05G6.9 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T29359  
R:Murray, J.; Le, T.T.  
submitted to the EMBL Data Library, May 1996  
A:Description: The sequence of C. elegans cosmid R05G6.  
A:Reference number: Z20612  
A:Accession: T29359  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-372 <MUR>  
A:Cross-references: EMBL:U58746; PIDN:AAR00626.1; GSPDB:GN00022; CESP:R05G6.9  
A:Experimental source: strain Bristol N2; clone R05G6  
C:Genetics:  
A:Gene: CESP:R05G6.9  
A:Map position: 4  
A:Introns: 80/1; 161/1; 245/1; 286/1

Query Match 75.9%; Score 41; DB 2; Length 372;  
Best Local Similarity 77.8%; Pred. No. 8.3;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9  
| | | | |  
Db 272 CAPGYVGR 280

RESULT 8

T27283  
hypothetical protein Y64G10A.f - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T27283  
R:Ainscough, R.  
submitted to the EMBL Data Library, September 1999  
A:Reference number: Z20336  
A:Accession: T27283  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1620 <WIL>  
A:Cross-references: EMBL:AL110498; NID:el542303; PIDN:CAB54471.1; CESP:Y64G10A.f  
A:Experimental source: clone Y64G10A  
C:Genetics:  
A:Gene: CESP:Y64G10A.f  
A:Introns: 77/1; 116/1; 198/1; 282/1; 365/1; 425/1; 466/1; 548/1; 559/1; 601/1; 625/1;

Query Match 75.9%; Score 41; DB 2; Length 1620;  
Best Local Similarity 66.7%; Pred. No. 36;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9  
| | | | |  
Db 1424 CPPGYIGTK 1432

RESULT 9

S23936  
N-selectin precursor - rat  
A:Alternate names: leucocyte cell adhesion molecule 1 (LECAM-1)  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 02-Dec-1993 #sequence\_revision 01-Sep-1995 #text\_change 20-Jun-2000  
C:Accession: S23936  
R:Watanabe, T.; Song, Y.; Hirayama, Y.; Tamatani, T.; Kuida, K.; Miyasaka, M.  
Biochim. Biophys. Acta 1131, 321-324, 1992  
A>Title: Sequence and expression of a rat cDNA for LECAM-1.

A:Reference number: S23936; MUID:92329548

A:Accession: S23936  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-372 <WAT>  
A:Cross-references: GB:D10831; NID:g220801; PIDN:BAA01613.1; PID:g220802  
C:Superfamily: L-selectin; C-type lectin homology; complement factor H repeat homology  
C:Keywords: transmembrane protein  
F:29-155/Domain: C-type lectin homology <LCH>  
F:160-191/Domain: EGF homology <EGF>  
F:197-254/Domain: complement factor H repeat homology <FHL>  
F:259-316/Domain: complement factor H repeat homology <FH2>

Query Match 74.1%; Score 40; DB 2; Length 372;  
Best Local Similarity 85.7%; Pred. No. 13;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDPGYIG 7  
| | | | |  
Db 182 CDPGYIG 188

RESULT 10

SI8188  
notch protein homolog - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 19-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 13-Aug-1999  
C:Accession: SI8188  
R:Weinmaster, G.; Roberts, V.J.; Lemke, G.  
Development 113, 199-205, 1991  
A>Title: A homolog of Drosophila Notch expressed during mammalian development.  
A:Reference number: SI8188; MUID:92111383  
A:Accession: SI8188  
A:Molecule type: mRNA  
A:Residues: 1-2531 <WET>  
A:Cross-references: EMBL:X57405; NID:g57634; PID:g57635  
C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homol  
F:987-1018/Domain: EGF homology <EGF1>  
F:1025-1056/Domain: EGF homology <EGF1>  
F:1233-1264/Domain: EGF homology <EGF2>  
F:1917-1949/Domain: ankyrin repeat homology <ANI>  
F:1950-1982/Domain: ankyrin repeat homology <ANI>  
F:1984-2016/Domain: ankyrin repeat homology <ANI>  
F:2017-2049/Domain: ankyrin repeat homology <ANI>  
F:2050-2082/Domain: ankyrin repeat homology <ANI>

Query Match 74.1%; Score 40; DB 2; Length 2531;  
Best Local Similarity 75.0%; Pred. No. 83;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDPGYIGS 8  
| | | | |  
Db 666 CEPGYTGS 673

RESULT 11

A46019  
Notch-1 protein - mouse  
N:Alternate names: notch protein  
C:Species: Mus musculus (house mouse)  
C>Date: 22-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Sep-1999  
C:Accession: A46019; S25144  
R:del Amo, F.F.; Gendron-Maguire, M.; Swiatek, P.J.; Jenkins, N.A.; Copeland, N.G.; G  
Genomics 15, 259-264, 1993  
A>Title: Cloning, analysis, and chromosomal localization of Notch-1, a mouse homolog  
A:Reference number: A46019; MUID:93194170  
A:Accession: A46019  
A:Status: not compared with conceptual translation  
A:Molecule type: nucleic acid  
A:Residues: 1-2531 <DEL>  
A:Cross-references: GB:Z11886; GB:S47228; NID:g288502; PIDN:CAA77941.1; PID:g288503

A:Note: sequence extracted from NCBI backbone (NCBIP:127318)  
R:Franco del Amo, F.; Smith, D.E.; Swiatek, P.J.; Gendron-Maguire, M.; Greenspan, R.J.;  
submitted to the EMBL Data Library, April 1992  
A:Description: Expression pattern of Notch, a mouse homolog of Drosophila Notch, suggest  
A:Reference number: S25144  
A:Accession: S25144  
A:Molecule type: mRNA  
A:Residues: 1551-2108, 'Q', 2110-2114, 'ALP', 2118-2170 <FRA>  
A:Cross-references: EMBL:Z11886  
C:Genetics:  
A:Gene: notch-1  
A:Map position: 2  
A:Note: proximal region of chromosome 2  
C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology  
F:106-138/Domain: EGF homology <EGF1>  
F:144-175/Domain: EGF homology <EG01>  
F:222-254/Domain: EGF homology <EGF2>  
F:261-292/Domain: EGF homology <EG02>  
F:339-370/Domain: EGF homology <EG03>  
F:416-449/Domain: EGF homology <EGF3>  
F:456-487/Domain: EGF homology <EG04>  
F:494-525/Domain: EGF homology <EG05>  
F:532-563/Domain: EGF homology <EG06>  
F:607-638/Domain: EGF homology <EG07>  
F:682-713/Domain: EGF homology <EG08>  
F:757-788/Domain: EGF homology <EG09>  
F:795-826/Domain: EGF homology <EG10>  
F:873-904/Domain: EGF homology <EG11>  
F:911-942/Domain: EGF homology <EG12>  
F:949-980/Domain: EGF homology <EG13>  
F:987-1018/Domain: EGF homology <EG14>  
F:1025-1056/Domain: EGF homology <EG15>  
F:1063-1094/Domain: EGF homology <EG16>  
F:1149-1180/Domain: EGF homology <EG17>  
F:1187-1218/Domain: EGF homology <EG18>  
F:1233-1264/Domain: EGF homology <EGF4>  
F:1352-1383/Domain: EGF homology <EG19>  
F:1391-1425/Domain: EGF homology <EGF>  
F:1917-1948/Domain: ankyrin repeat homology <AN1>  
F:1949-1981/Domain: ankyrin repeat homology <AN2>  
F:1983-2015/Domain: ankyrin repeat homology <AN3>  
F:2016-2048/Domain: ankyrin repeat homology <AN4>  
F:2049-2081/Domain: ankyrin repeat homology <AN5>

Query Match 74.1%; Score 40; DB 2; Length 2531;  
Best Local Similarity 75.0%; Pred. No. 83;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDPGYIGS 8  
|:|:|:|:|  
Db 666 CEPGYTGS 673

RESULT 12

A40043  
notch protein homolog TAN-1 precursor - human  
C:Species: Homo sapiens (man)  
C:Date: 21-Apr-1992 #sequence\_revision 21-Apr-1992 #text\_change 13-Aug-1999  
C:Accession: A40043  
R:Ellisen, L.W.; Bird, J.; West, D.C.; Soreng, A.L.; Reynolds, T.C.; Smith, S.D.; Sklar,  
Cell 66, 649-661, 1991  
A:Title: TAN-1, the human homolog of the Drosophila Notch gene, is broken by chromosomal  
A:Reference number: A40043; MUID:91347367  
A:Accession: A40043  
A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra  
A:Molecule type: mRNA  
A:Residues: 1-2555 <ELL>  
A:Cross-references: GB:W73980  
C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology  
F:261-292/Domain: EGF homology <EGX1>  
F:494-525/Domain: EGF homology <EGF1>  
F:987-1018/Domain: EGF homology <EGX2>

F:1149-1180/Domain: EGF homology <EGF>  
F:1187-1218/Domain: EGF homology <EGF3>  
F:1233-1264/Domain: EGF homology <EGX3>  
F:1927-1959/Domain: ankyrin repeat homology <AN1>  
F:1960-1992/Domain: ankyrin repeat homology <AN2>  
F:1994-2026/Domain: ankyrin repeat homology <AN3>  
F:2027-2059/Domain: ankyrin repeat homology <AN4>  
F:2060-2092/Domain: ankyrin repeat homology <AN5>

Query Match 74.1%; Score 40; DB 2; Length 2555;  
Best Local Similarity 75.0%; Pred. No. 84;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDPGYIGS 8  
|:|:|:|:|  
Db 665 CEPGYTGS 672

RESULT 13

A24420  
notch protein - fruit fly (Drosophila melanogaster)  
N:Alternate names: neurogenic repetitive locus protein  
C:Species: Drosophila melanogaster  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: A24420; A24768; S09358; A05267  
R:Kidd, S.; Kelley, M.R.; Young, M.W.  
Mol. Cell. Biol. 6, 3094-3108, 1986  
A:Reference number: A24420; MUID:87064624  
A:Accession: A24420  
A:Molecule type: DNA  
A:Residues: 1-2703 <KID>  
A:Cross-references: GB:K03508; NID:g157991; PIDN:AAA28725.1; PID:g157993  
R:Wharton, K.A.; Johansen, K.M.; Xu, T.; Artavanis-Tsakonas, S.  
Cell 43, 567-581, 1985  
A:Reference number: A24768; MUID:86079539  
A:Accession: A24768  
A:Molecule type: mRNA  
A:Residues: 1-48, 'I', 50-118, 'R', 120-230, 'I', 232-256, 'N', 258-266, 'A', 268-872, 'R', 874-9  
A:Note: the authors translated the codon ATC for residue 49 as Thr, ATT for residue 2  
R:Tautz, D.  
Nucleic Acids Res. 17, 6463-6471, 1989

A:Title: Hypervariability of simple sequences as a general source for polymorphic DNA  
A:Reference number: S09358; MUID:89385974  
A:Accession: S09358

A:Molecule type: DNA  
A:Residues: 2505-2551, 'QQQQ', 2552-2576, 'E', 2578-2604 <TAU>  
R:Wharton, K.A.; Yedvobnick, B.; Finnerty, V.G.; Artavanis-Tsakonas, S.  
Cell 40, 55-62, 1985  
A:Title: opa: a novel family of transcribed repeats shared by the Notch locus and oth  
A:Reference number: A05267; MUID:85099329  
A:Accession: A05267  
A:Molecule type: DNA  
A:Residues: 2504-2576, 'E', 2578-2611 <WHA2>  
C:Genetics:

A:Gene: notch; opa  
A:Cross-references: FlyBase:FBgn0004647  
A:Map position: 8.96-9.36  
A:Introns: 53/3; 84/3; 171/3; 240/3; 283/3; 2333/3; 2436/3; 2588/3  
C:Superfamily: notch protein; ankyrin repeat homology; EGF homology  
C:Keywords: differentiation; tandem repeat; transmembrane protein  
F:27-43/Domain: transmembrane #status predicted <TMW1>  
F:297-328/Domain: EGF homology <EGX1>  
F:530-561/Domain: EGF homology <EGF1>  
F:568-599/Domain: EGF homology <EGF>  
F:988-1019/Domain: EGF homology <EGX2>  
F:1064-1095/Domain: EGF homology <EGF3>  
F:1187-1218/Domain: EGF homology <EGX3>  
F:1746-1762/Domain: transmembrane #status predicted <TMW2>  
F:1950-1982/Domain: ankyrin repeat homology <AN1>  
F:1983-2015/Domain: ankyrin repeat homology <AN2>  
F:1988-2004/Domain: transmembrane #status predicted <TMW3>  
F:2017-2049/Domain: ankyrin repeat homology <AN3>

F:2050-2082/Domain: ankyrin repeat homology <AN4>  
 F:2083-2115/Domain: ankyrin repeat homology <AN5>  
 F:2538-2568/Region: glutamine-rich  
 F:2538-2568/Domain: neurogenic repetitive element #status predicted <OPA>

Query Match 74.1%; Score 40; DB 1; Length 2703;  
 Best Local Similarity 85.7%; Pred. No. 89;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDPGYIG 7  
 |||||  
 Db 627 CDPGYTG 633

## RESULT 14

A40701  
 tenascin-X precursor - human  
 C:Species: Homo sapiens (man)  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Dec-1999  
 C:Accession: A40701; A33725; C42175  
 R:Bristow, J.; Tee, M.K.; Gitelman, S.E.; Mellon, S.H.; Miller, W.L.  
 J. Cell Biol. 122, 265-278, 1993  
 A:Title: Tenascin-X: a novel extracellular matrix protein encoded by the human XB gene  
 A:Reference number: A40701; MUID:93300909  
 A:Accession: A40701  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-3566 <BRI>  
 A:Cross-references: EMBL:X71937  
 R:Morel, Y.; Bristow, J.; Gitelman, S.E.; Miller, W.L.  
 Proc. Natl. Acad. Sci. U.S.A. 86, 6582-6586, 1989  
 A:Title: Transcript encoded on the opposite strand of the human steroid 21-hydroxylase/  
 A:Reference number: A33725; MUID:89367293  
 A:Accession: A33725  
 A:Molecule type: mRNA  
 A:Residues: 2748-3199, 'Y', 3201-3298, 'E', 3299-3314, 'G', 3316-3566 <MOR>  
 R:Matsumoto, K.; Aral, M.; Ishihara, N.; Ando, A.; Inoko, H.; Ikemura, T.  
 Genomics 12, 485-491, 1992  
 A:Title: Cluster of fibronectin type III repeats found in the human major histocompatibility  
 enascin.  
 A:Reference number: A42175; MUID:92217969  
 A:Accession: C42175  
 A:Molecule type: DNA  
 A:Residues: 1849-1936 <MAT>  
 A:Experimental source: clone 3.9kF3-1  
 A:Note: sequence extracted from NCBI backbone (NCBIP:95694)  
 A:Gene: GDB:TNXA; D6S103E; TNX; XA; XB  
 A:Cross-references: GDB:568487; OMIM:600261  
 A:Map position: 6p21.3-6p21.3  
 C:Superfamily: tenascin-X; EGF homology; fibrinogen beta/gamma homology; fibronectin type  
 C:Keywords: extracellular matrix; glycoprotein  
 F:435-461/Domain: EGF homology <EGF>  
 F:748-828/Domain: fibrinogen beta/gamma homology; fibronectin type  
 F:829-856/Domain: fibronectin type III repeat homology <3F1>  
 F:873-953/Domain: fibronectin type III repeat homology #status atypical <3F1>  
 F:975-1055/Domain: fibronectin type III repeat homology <3F3>  
 F:1078-1158/Domain: fibronectin type III repeat homology <3F4>  
 F:1167-1247/Domain: fibronectin type III repeat homology <3F5>  
 F:1248-1317/Domain: fibronectin type III repeat homology <3F6>  
 F:1323-1403/Domain: fibronectin type III repeat homology #status atypical <3F7>  
 F:1412-1492/Domain: fibronectin type III repeat homology <3F8>  
 F:1510-1590/Domain: fibronectin type III repeat homology <3F9>  
 F:1618-1676/Domain: fibronectin type III repeat homology <3F10>  
 F:1678-1749/Domain: fibronectin type III repeat homology #status atypical <3F11>  
 F:1753-1831/Domain: fibronectin type III repeat homology #status atypical <3F12>  
 F:1849-1929/Domain: fibronectin type III repeat homology <3F13>  
 F:1955-2035/Domain: fibronectin type III repeat homology <3F14>  
 F:2061-2141/Domain: fibronectin type III repeat homology <3F15>  
 F:2167-2246/Domain: fibronectin type III repeat homology <3F16>  
 F:2274-2354/Domain: fibronectin type III repeat homology <3F17>

F:2382-2462/Domain: fibronectin type III repeat homology <3F19>  
 F:2488-2568/Domain: fibronectin type III repeat homology <3F20>  
 F:2584-2664/Domain: fibronectin type III repeat homology <3F21>  
 F:2677-2757/Domain: fibronectin type III repeat homology <3F22>  
 F:2771-2851/Domain: fibronectin type III repeat homology <3F23>  
 F:2878-2958/Domain: fibronectin type III repeat homology <3F24>  
 F:2977-3067/Domain: fibronectin type III repeat homology #status atypical <3F25>  
 F:3078-3159/Domain: fibronectin type III repeat homology <3F26>  
 F:3167-3247/Domain: fibronectin type III repeat homology <3F27>  
 F:3255-3334/Domain: fibronectin type III repeat homology <3F28>  
 F:3349-3557/Domain: fibrinogen beta/gamma homology <FBG>

Query Match 74.1%; Score 40; DB 1; Length 3566;  
 Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDPGYIG 7  
 |||||  
 Db 266 CDPGYTG 272

## RESULT 15

T42629  
 tenascin-X - bovine  
 N:Alternate names: flexillin  
 C:Species: Bos primigenius taurus (cattle)  
 C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 21-Jul-2000  
 C:Accession: T42629  
 R:Eleftheriou, F.; Exposito, J.Y.; Garrone, R.; Lethias, C.  
 J. Biol. Chem. 272, 22866-22874, 1997  
 A:Title: Characterization of the bovine tenascin-X.  
 A:Reference number: 22180; MUID:97426436  
 A:Accession: T42629  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-4135 <ELE>  
 A:Cross-references: EMBL:Y11915; NID:g2462978; PIDN:CAA72671.1; PID:g2462979  
 C:Genetics:  
 A:Gene: TN-X  
 C:Superfamily: tenascin-X; EGF homology; fibrinogen beta/gamma homology; fibronectin  
 C:Keywords: extracellular matrix; glycoprotein; heptad repeat

Query Match 74.1%; Score 40; DB 2; Length 4135;  
 Best Local Similarity 85.7%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDPGYIG 7  
 |||||  
 Db 327 CDPGYTG 333

Search completed: August 15, 2001, 10:54:17  
 Job time: 170 sec

---

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2001, 10:55:54 ; Search time 12.86 Seconds  
(without alignments)  
23.974 Million cell updates/sec

Title: US-09-673-785A-4

Perfect score: 54

Sequence: 1 CDPGYIGSR 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54	100.0	1786	1 LMB1_HUMAN	P07942 homo sapien
2	54	100.0	1786	1 LMB1_MOUSE	P02469 mus musculus
3	43	79.6	178	1 B7C_HUMAN	P35070 homo sapien
4	43	79.6	655	1 ITB5_PAPCY	Q07441 papio cynoc
5	43	79.6	799	1 ITB5_HUMAN	P18084 homo sapien
6	40	74.1	372	1 LEM1_RAT	P18036 rattus norv
7	40	74.1	618	1 DLL3_HUMAN	Q90VJ7 homo sapien
8	40	74.1	2444	1 NTC1_HUMAN	P46531 homo sapien
9	40	74.1	2531	1 NTC1_MOUSE	Q01705 mus musculus
10	40	74.1	2531	1 NTC1_RAT	Q07008 rattus norv
11	40	74.1	2703	1 NOTC_DROME	P07207 drosophila
12	40	74.1	4289	1 TENX_HUMAN	P22105 homo sapien
13	39	72.2	810	1 NELL_RAT	Q62919 rattus norv
14	39	72.2	1295	1 GLP1_CAEL	P13508 caenorhabdi
15	39	72.2	1964	1 NTC4_MOUSE	P31695 mus musculus
16	39	72.2	2437	1 NOTC_BRARE	P46530 brachydanio
17	38	70.4	227	1 YMQ4_CAEL	P34495 caenorhabdi
18	38	70.4	372	1 CNTR_HUMAN	P26992 homo sapien
19	38	70.4	810	1 NELL_HUMAN	Q92832 homo sapien
20	38	70.4	955	1 TSP4_XENLA	Q06441 xenopus lae
21	38	70.4	1480	1 SLIT_DROME	P24014 drosophila
22	38	70.4	2481	1 UN52_CAEL	Q08561 caenorhabdi
23	38	70.4	2524	1 NOTC_XENLA	P21783 xenopus lae
24	38	70.4	3672	1 LML2_CAEL	Q21313 caenorhabdi
25	38	70.4	4393	1 PCBM_HUMAN	P98160 homo sapien
26	37	68.5	298	1 WTE5_ECOLI	P04393 escherichia
27	37	68.5	589	1 DLL3_RAT	O88671 rattus norv
28	37	68.5	592	1 DLL3_MOUSE	O88516 mus musculus
29	37	68.5	769	1 ITB2_BOVIN	P32592 bos taurus
30	37	68.5	769	1 ITB2_HUMAN	P05107 homo sapien
31	37	68.5	769	1 ITB2_PIG	P53714 sus scrofa
32	37	68.5	956	1 TSP3_MOUSE	Q03895 mus musculus
33	37	68.5	1025	1 CR2_MOUSE	P19070 mus musculus

34	37	68.5	2039	1 CRL_HUMAN	P17927 homo sapien
35	37	68.5	3110	1 LMA2_HUMAN	P24043 homo sapien
36	37	68.5	3707	1 PGBM_MOUSE	Q05793 mus musculus
37	37	68.5	3712	1 LMA_DROME	Q00174 drosophila
38	36	66.7	127	1 REGO_ECOLI	Q47274 escherichia
39	36	66.7	208	1 HBGF_CERAE	Q09118 cercopithec
40	36	66.7	208	1 HBGF_HUMAN	Q99075 homo sapien
41	36	66.7	208	1 HBGF_PIG	P50396 sus scrofa
42	36	66.7	323	1 GDIA_MOUSE	Q09671 schizosacch
43	36	66.7	392	1 OYER_SCHPO	P22891 homo sapien
44	36	66.7	400	1 PRT2_HUMAN	P31150 homo sapien
45	36	66.7	447	1 GDIA_HUMAN	

#### ALIGNMENTS

```

RESULT 1
LMB1_HUMAN
ID LMB1_HUMAN STANDARD; PRT; 1786 AA.
AC P07942;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE LAMININ BETA-1 CHAIN PRECURSOR (LAMININ B1 CHAIN).
GN LAMB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90368768; PubMed=1975589;
RA Vuolteenaho R., Chow L.T., Tryggvason K.;
RT "Structure of the human laminin B1 chain gene.";
RL J. Biol. Chem. 265:15611-15616(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87280097; PubMed=3611077;
RA Pikkarainen T., Eddy R., Fukushima Y., Byers M., Shows T.,
RA Pihlajaniemi T., Saraste M., Tryggvason K.;
RT "Human laminin B1 chain. A multidomain protein with gene (LAMB1)
RL locus in the q22 region of chromosome 7.";
RN [3]
RX MEDLINE=88021029; PubMed=3661559;
RA Jaye M., Modi W.S., Ricca G.A., Mudd R., Chiu I.M., O'Brien S.J.,
RA Drohan W.N.;
RT "Isolation of a cDNA clone for the human laminin-B1 chain and its
RL gene localization.";
RN [4]
RX MEDLINE=41605615(1987).
RA Am. J. Hum. Genet. 41:605-615(1987).
CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
CC CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
CC -! SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
CC DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
CC TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
CC COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
CC THE BETA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ), LAMININ-
CC 2 (MEROSIN), AND LAMININ-6 (K-LAMININ).
CC -! SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -! TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
CC COMPONENT).
CC -! DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -! DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
CC -! SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -! SIMILARITY: CONTAINS 12.5 LAMININ EGF-LIKE DOMAINS.
CC -! SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.

```

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration



Query Match 100.08; Score 54; DB 1; Length 1786;  
 Best Local Similarity 100.0%; Pred. No. 0.064;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9  
 |||||  
 Db 946 CDPGYIGSR 954

## RESULT 2

LMBL\_MOUSE STANDARD; PRT: 1786 AA.  
 AC P02469;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE LAMININ BETA-1 CHAIN PRECURSOR (LAMININ B1 CHAIN).  
 GN LMBL-1 OR LAMB-1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 ON NCBI\_TaxID=10090;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-8714712; PubMed-3493487;  
 RA Sasaki M., Kato S., Kohno K., Martin G.R., Yamada Y.;  
 RA Barlow D.P., Green N.M., Kurkinen M., Hogan B.L.M.;  
 RT "Sequencing of laminin B chain cDNAs reveals C-terminal regions of  
 coiled-coil alpha-helix";  
 RL EMBO J. 3:2355-2362(1984).  
 [3]  
 RP SEQUENCE OF 165-172; 539-547 AND 712-719.  
 RX STRAIN-BALB/C; TISSUE-Endothelial cells;  
 RC MEDLINE-97363207; PubMed-9219532;  
 RA Frieser M., Noeckel H., Pausch F., Roeder C., Hahn A., Deutzmann R.,  
 RA Sorokin L.M.;  
 RT "Cloning of the mouse laminin alpha 4 cDNA. Expression in a subset of  
 endothelium";  
 RL Eur. J. Biochem. 246:727-735(1997).  
 CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR. LAMININ  
 IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF  
 CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING  
 WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.  
 CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE  
 DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND  
 TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE  
 COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.  
 CC THE BETA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ), LAMININ-  
 2 (MEROSIN), AND LAMININ-6 (K-LAMININ).  
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
 CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR  
 COMPONENT).  
 CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).  
 CC -1- SIMILARITY: CONTAINS 12.5 LAMININ EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; M15525; AAA39407.1; ALT\_INIT.  
 CC EMBL; X05212; CAA28839.1; -.  
 DR

PIR: A26413; MMSB1.  
 HSP; P03069; LZIM.  
 DR MGD; MGI:96743; Lamb1-1.  
 DR InterPro; IPR000561; -.  
 DR InterPro; IPR001886; -.  
 DR InterPro; IPR002049; -.  
 DR Pfam; PF00053; laminin\_EGF; 13.  
 DR Pfam; PF00055; laminin\_Nterm; 1.  
 DR PRINTS; PRO0011; EGF\_LAMININ.  
 DR PROSITE; PS00022; EGF\_1; 9.  
 DR PROSITE; PS01186; EGF\_2; 2.  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 11.  
 KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;  
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.  
 FT SIGNAL 1 21  
 FT CHAIN 22 1786 LAMININ BETA-1 CHAIN.  
 FT DOMAIN 22 270 LAMININ N-TERMINAL (DOMAIN VI).  
 FT DOMAIN 271 540 4.5 X LAMININ EGF-LIKE REPEATS (DOMAIN  
 V).  
 FT DOMAIN 271 334 LAMININ EGF-LIKE 1.  
 FT DOMAIN 335 397 LAMININ EGF-LIKE 2.  
 FT DOMAIN 398 457 LAMININ EGF-LIKE 3.  
 FT DOMAIN 458 509 LAMININ EGF-LIKE 4.  
 FT DOMAIN 510 540 LAMININ EGF-LIKE 5 (INCOMPLETE).  
 FT DOMAIN 541 772 LAMININ DOMAIN IV.  
 FT DOMAIN 773 1178 8 X LAMININ EGF-LIKE REPEATS (DOMAIN  
 III).  
 FT DOMAIN 773 820 LAMININ EGF-LIKE 6.  
 FT DOMAIN 821 866 LAMININ EGF-LIKE 7.  
 FT DOMAIN 867 916 LAMININ EGF-LIKE 8.  
 FT DOMAIN 917 975 LAMININ EGF-LIKE 9.  
 FT DOMAIN 976 1027 LAMININ EGF-LIKE 10.  
 FT DOMAIN 1028 1083 LAMININ EGF-LIKE 11.  
 FT DOMAIN 1084 1131 LAMININ EGF-LIKE 12.  
 FT DOMAIN 1132 1178 LAMININ EGF-LIKE 13.  
 FT DOMAIN 1179 1397 DOMAIN II.  
 FT DOMAIN 1398 1430 DOMAIN ALPHA.  
 FT DOMAIN 1431 1786 DOMAIN I.  
 FT DOMAIN 1216 1315 COILED COIL (POTENTIAL).  
 FT DOMAIN 1368 1388 COILED COIL (POTENTIAL).  
 FT DOMAIN 1448 1778 COILED COIL (POTENTIAL).  
 FT DISULFID 271 280 BY SIMILARITY.  
 FT DISULFID 273 288 BY SIMILARITY.  
 FT DISULFID 300 309 BY SIMILARITY.  
 FT DISULFID 312 332 BY SIMILARITY.  
 FT DISULFID 335 344 BY SIMILARITY.  
 FT DISULFID 337 362 BY SIMILARITY.  
 FT DISULFID 365 374 BY SIMILARITY.  
 FT DISULFID 377 395 BY SIMILARITY.  
 FT DISULFID 398 411 BY SIMILARITY.  
 FT DISULFID 400 426 BY SIMILARITY.  
 FT DISULFID 428 437 BY SIMILARITY.  
 FT DISULFID 440 455 BY SIMILARITY.  
 FT DISULFID 458 472 BY SIMILARITY.  
 FT DISULFID 460 479 BY SIMILARITY.  
 FT DISULFID 481 490 BY SIMILARITY.  
 FT DISULFID 493 507 BY SIMILARITY.  
 FT DISULFID 773 785 BY SIMILARITY.  
 FT DISULFID 775 792 BY SIMILARITY.  
 FT DISULFID 794 803 BY SIMILARITY.  
 FT DISULFID 806 818 BY SIMILARITY.  
 FT DISULFID 821 833 BY SIMILARITY.  
 FT DISULFID 823 840 BY SIMILARITY.  
 FT DISULFID 842 851 BY SIMILARITY.  
 FT DISULFID 854 864 BY SIMILARITY.  
 FT DISULFID 867 876 BY SIMILARITY.  
 FT DISULFID 869 883 BY SIMILARITY.  
 FT DISULFID 886 895 BY SIMILARITY.  
 FT DISULFID 917 933 BY SIMILARITY.  
 FT DISULFID 919 944 BY SIMILARITY.  
 FT DISULFID 946 955 BY SIMILARITY.  
 FT DISULFID 958 973 BY SIMILARITY.

FT DISULFID 976 990 BY SIMILARITY.  
 FT DISULFID 978 997 BY SIMILARITY.  
 FT DISULFID 1000 1009 BY SIMILARITY.  
 FT DISULFID 1012 1025 BY SIMILARITY.  
 FT DISULFID 1084 1096 BY SIMILARITY.  
 FT DISULFID 1086 1103 BY SIMILARITY.  
 FT DISULFID 1105 1114 BY SIMILARITY.  
 FT DISULFID 1117 1129 BY SIMILARITY.  
 FT DISULFID 1132 1144 BY SIMILARITY.  
 FT DISULFID 1134 1151 BY SIMILARITY.  
 FT DISULFID 1153 1162 BY SIMILARITY.  
 FT DISULFID 1155 1176 BY SIMILARITY.  
 FT DISULFID 1179 1179 INTERCHAIN (PROBABLE).  
 FT DISULFID 1182 1182 INTERCHAIN (PROBABLE).  
 FT DISULFID 1785 1785 INTERCHAIN (PROBABLE).  
 FT CARBOHYD 120 120 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 677 677 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1041 1041 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1195 1195 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1279 1279 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1336 1336 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1487 1487 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1533 1533 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1542 1542 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1643 1643 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 1531 1534 SGNA -> MEMP (IN REF. 2).  
 FT CONFLICT 1749 1749 D -> N (IN REF. 2).  
 SQ SEQUENCE 1786 AA; 196904 MW; 846671B7BF41474 CRC64;

Query Match 100.0%; Score 54; DB 1; Length 1786;  
 Best Local Similarity 100.0%; Pred. No. 0.064;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9  
 |||||  
 Db 946 CDPGYIGSR 954

RESULT 3  
 BTC\_HUMAN  
 ID BTC\_HUMAN STANDARD; PRT; 178 AA.  
 AC P35070;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE BETACELLULIN PRECURSOR (BTC).  
 GN BTC.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Breast;  
 RX MEDLINE=93176165; PubMed=8439318;  
 RA Sasada R., Ono Y., Taniyama Y., Shing Y., Folkman J., Igarashi K.;  
 RT "Cloning and expression of cDNA encoding human betacellulin, a new  
 RL member of the EGF family.";  
 RL Biochem. Biophys. Res. Commun. 190:1173-1179(1993).  
 CC -!- FUNCTION: POTENT MITOGEN FOR RETINAL PIGMENT EPITHELIAL CELLS  
 CC AND VASCULAR SMOOTH MUSCLE CELLS. THE EFFECTS OF BETACELLULIN  
 CC ARE PROBABLY MEDIATED BY THE EGF RECEPTOR AND OTHER RELATED  
 CC RECEPTORS.  
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (PRECURSOR FORM);  
 CC EXTRACELLULAR (MATURE FORM).  
 CC -!- TISSUE SPECIFICITY: SYNTHESIZED IN SEVERAL TISSUES AND TUMOR  
 CC CELLS. PROBABLY NOT FOUND IN THE BRAIN.  
 CC -!- DISEASE: BETACELLULIN FROM BETA CELLS COULD PLAY A ROLE IN THE  
 CC VASCULAR COMPLICATIONS ASSOCIATED WITH DIABETES.

CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.  
 CC -!- SIMILARITY: STRONG, TO MOUSE BETACELLULIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; S55606; AAB25452.1;  
 DR PIR; JCI467; JCI467.  
 DR HSP; F01135; IYUF.  
 DR MIM; 600345;  
 DR InterPro; IPR000561;  
 DR InterPro; IPR001336;  
 DR Pfam; PF00008; EGF; 1.  
 DR PRINTS; PR00009; EGF\_TGF.  
 DR PROSITE; PS00022; EGF\_1; 1.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 KW Growth factor; Mitogen; Glycoprotein; EGF-like domain; Transmembrane;  
 KW Signal.  
 FT SIGNAL 1 31 POTENTIAL.  
 FT CHAIN 32 111 BETACELLULIN.  
 FT PROPEP 112 178 REMOVED IN MATURE FORM.  
 FT DOMAIN 32 118 EXTRACELLULAR (POTENTIAL).  
 FT TRANSEM 119 139 POTENTIAL.  
 FT DOMAIN 140 178 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 146 154 ARG/LYS-RICH (BASIC).  
 FT CARBOHYD 34 34 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT DOMAIN 65 105 EGF-LIKE.  
 FT DISULFID 69 82 BY SIMILARITY.  
 FT DISULFID 77 93 BY SIMILARITY.  
 FT DISULFID 95 104 BY SIMILARITY.  
 SQ SEQUENCE 178 AA; 19746 MW; 27AC77BD92001F0F CRC64;

Query Match 79.6%; Score 43; DB 1; Length 178;  
 Best Local Similarity 77.8%; Pred. No. 0.61;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9  
 |||||  
 Db 95 CDEGYIGAR 103

RESULT 4  
 ITB5\_PAPCY  
 ID ITB5\_PAPCY STANDARD; PRT; 655 AA.  
 AC Q07441;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE INTEGRIN BETA-5 (FRAGMENT).  
 GN ITGB5.  
 OS Papio cynocephalus (Yellow baboon).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;  
 OX NCBI\_TaxID=9556;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94040831; PubMed=8224922;  
 RA Shoji M., Hayzer D.J., Kim T.M., Runge M.S., Hanson S.R.;  
 RT "Human and baboon integrin beta 5 subunit-encoding mRNAs have  
 RL alternative polyadenylation sites.";  
 RL Gene 133:307-308(1993).  
 CC -!- FUNCTION: INTEGRINS ARE A LARGE FAMILY OF CELL SURFACE  
 CC GLYCOPROTEINS THAT MEDIATE CELL TO CELL & CELL TO MATRIX ADHESION.  
 CC -!- SUBUNIT: DIMER OF AN ALPHA AND BETA SUBUNIT. BETA-5 ASSOCIATES  
 CC WITH ALPHA-V.  
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.



Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 CDPGYIGSR 9  
Db 489 CSPGYLGR 497

## RESULT 6

LEML\_RAT  
ID LEML\_RAT STANDARD; PRT; 372 AA.  
AC P30836;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-FEB-1996 (Rel. 26, Last sequence update)  
DE L-SELECTIN PRECURSOR (LYMPH NODE HOMING RECEPTOR)  
DE MOLECULE-1 (LAW-1) (LY-22) (LYMPHOCYTE SURFACE MEL-14 ANTIGEN)  
DE (LEUKOCYTE-ENDOTHELIAL CELL ADHESION MOLECULE 1) (LECAM1) (CD62L).  
OS SELL OR LNHR OR LY-22.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
RN [1]  
RP MEDLINE-92329548; PubMed-1378303;  
RA Watanabe T., Song Y., Hirayama Y., Tamatani T., Kuida K., Miyasaka M.;  
RT "Sequence and expression of a rat cDNA for LECAM-1.";  
RL Blochim. Biophys. Acta 1131:321-324(1992).  
CC -|- FUNCTION: CELL SURFACE ADHESION PROTEIN. MEDIATES THE ADHERENCE  
CC OF LYMPHOCYTES TO ENDOTHELIAL CELLS OF HIGH ENDOTHELIAL  
CC VENULES IN PERIPHERAL LYMPH NODES.  
CC -|- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -|- SIMILARITY: TO OTHER SELECTINS/LECAMs.  
CC -|- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
CC -|- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.  
CC -|- SIMILARITY: CONTAINS 2 SUSHI (SCR) REPEATS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
-----  
EMBL: D10831; BAA01613.1;  
DR PBL; S23936; S23936.  
DR HSP; P14151; 1KJB.  
DR InterPro: IPR000436;  
DR InterPro: IPR000561;  
DR InterPro: IPR001304;  
DR InterPro: IPR002396;  
DR Pfam: PF00008; EGF; 1.  
DR Pfam: PF00059; lectin\_C; 1.  
DR Pfam: PF00084; sushi; 2.  
DR PRINTS: PR00343; SELECTIN.  
DR PROSITE: PS00022; EGF\_1; 1.  
DR PROSITE: PS01186; EGF\_2; 1.  
DR PROSITE: PS00615; C-TYPE\_LECTIN\_1; 1.  
DR PROSITE: PS50041; C-TYPE\_LECTIN\_2; 1.  
KW Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;  
KW Selectin; Signal; Sushi; Repeat.  
FT SIGNAL 1 28  
FT PROPEP 29 38 BY SIMILARITY.  
FT CHAIN 39 372 L-SELECTIN.  
FT DOMAIN 39 332 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 333 372 POTENTIAL.  
FT DOMAIN 356 372 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 55 155 C-TYPE LECTIN (SHORT FORM).  
FT DOMAIN 156 192 EGF-LIKE.  
FT DOMAIN 196 317 2 X SUSHI (SCR) REPEATS.  
FT DOMAIN 196 255 SUSHI 1.  
FT DOMAIN 258 317 SUSHI 2.

FT DISULFID 57 155 BY SIMILARITY.  
FT DISULFID 128 147 BY SIMILARITY.  
FT DISULFID 160 171 BY SIMILARITY.  
FT DISULFID 165 180 BY SIMILARITY.  
FT DISULFID 182 191 BY SIMILARITY.  
FT DISULFID 197 241 BY SIMILARITY.  
FT DISULFID 227 254 BY SIMILARITY.  
FT DISULFID 259 303 BY SIMILARITY.  
FT DISULFID 289 316 BY SIMILARITY.  
FT CARBOHYD 60 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 226 226 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 278 278 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 372 AA; 42441 MW; 3B88AE0F1E4D191A CRC64;

Query Match 74.1%; Score 40; DB 1; Length 372;  
Best Local Similarity 85.7%; Pred. No. 4.5;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 CDPGYIG 7  
Db 182 CDPGYIG 188

## RESULT 7

DL3\_HUMAN  
ID DL3\_HUMAN STANDARD; PRT; 618 AA.  
AC Q9NYJ7;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DE DELTA-LIKE PROTEIN 3 PRECURSOR (DROSOPHILA DELTA HOMOLOG 3).  
GN DL3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.; AND VARIANT SD ASP-385.  
RX MEDLINE-20206573; PubMed-10742114;  
RA Bulman M.P., Kusumi K., Frayling T.M., McKeown C., Garrett C.,  
RA Lander E.S., Krumlauf R., Hattersley A.T., Ellard S., Turpenny P.D.;  
RT "Mutations in the human delta homologue, DL3, cause axial skeletal  
RT defects in spondylocostal dysostosis.";  
RL Nat. Genet. 24:438-441(2000).  
CC -|- FUNCTION: INHIBITS PRIMARY NEUROGENESIS. MAY BE REQUIRED TO DIVERT  
CC NEURONS ALONG A SPECIFIC DIFFERENTIATION PATHWAY. PLAY A ROLE IN  
CC THE FORMATION OF SOMITE BOUNDARIES DURING SEGMENTATION OF THE  
CC PARAXIAL MESODERM (BY SIMILARITY).  
CC -|- SUBUNIT: CAN BIND AND ACTIVATE NOTCH-1 OR ANOTHER NOTCH RECEPTOR  
CC (BY SIMILARITY).  
CC -|- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (PROBABLE).  
CC -|- DOMAIN: THE DELTA-SERRATE-LAG2 (DSL) DOMAIN IS REQUIRED FOR  
CC BINDING TO THE NOTCH RECEPTOR.  
CC -|- DISEASE: DEFECTS IN DL3 ARE A CAUSE OF AUTOSOMAL RECESSIVE  
CC SPONDYLOCOSTAL DYSOSTOSIS (SD). IT IS CHARACTERIZED BY MULTIPLE  
CC HEMIVERTERAE, RIB FUSIONS AND DELETIONS WITH A NON-PROGRESSIVE  
CC KYPHOSCOLIOSIS.  
CC -|- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.  
CC -|- SIMILARITY: BELONGS TO THE DELTA/SERRATE/JAGGED FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
-----  
EMBL: AF241373; AAF62542.1;  
DR

DR EMBL; AF241367; AAF62542.1; JOINED.  
 DR EMBL; AF241368; AAF62542.1; JOINED.  
 DR EMBL; AF241369; AAF62542.1; JOINED.  
 DR EMBL; AF241370; AAF62542.1; JOINED.  
 DR EMBL; AF241371; AAF62542.1; JOINED.  
 DR EMBL; AF241372; AAF62542.1; JOINED.  
 DR MIM; 602768; -.  
 DR MIM; 277300; -.  
 DR InterPro; IPR000561; -.  
 DR InterPro; IPR000742; -.  
 DR InterPro; IPR001438; -.  
 DR PROSITE; PS00022; EGF\_1; 6.  
 DR PROSITE; PS01186; EGF\_2; 6.  
 KW Signal; EGF-like domain; Repeat; Transmembrane; Developmental protein;  
 KW Differentiation; Disease mutation.  
 FT SIGNAL 1 26  
 FT CHAIN 27 618 DELTA-LIKE PROTEIN 3.  
 FT DOMAIN 27 492 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 493 513 POTENTIAL.  
 FT DOMAIN 514 618 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 176 215 DELTA-SERRATE-LAG2.  
 FT DOMAIN 237 248 EGF-LIKE 1.  
 FT DOMAIN 298 309 EGF-LIKE 2.  
 FT DOMAIN 339 350 EGF-LIKE 3.  
 FT DOMAIN 377 388 EGF-LIKE 4.  
 FT DOMAIN 415 426 EGF-LIKE 5.  
 FT DOMAIN 453 464 EGF-LIKE 6.  
 FT DISULFID 239 248 BY SIMILARITY.  
 FT DISULFID 283 298 BY SIMILARITY.  
 FT DISULFID 300 309 BY SIMILARITY.  
 FT DISULFID 327 339 BY SIMILARITY.  
 FT DISULFID 341 350 BY SIMILARITY.  
 FT DISULFID 357 368 BY SIMILARITY.  
 FT DISULFID 362 377 BY SIMILARITY.  
 FT DISULFID 379 388 BY SIMILARITY.  
 FT DISULFID 395 406 BY SIMILARITY.  
 FT DISULFID 400 415 BY SIMILARITY.  
 FT DISULFID 417 426 BY SIMILARITY.  
 FT DISULFID 433 444 BY SIMILARITY.  
 FT DISULFID 438 453 BY SIMILARITY.  
 FT DISULFID 455 464 BY SIMILARITY.  
 FT VARIANT 385 385 G -> D (IN SD).  
 FT SEQUENCE 618 AA; 64617 MW; 58A9BC0A7DEAD1A0 CRC64;  
 Query Match 74.1%; Score 40; DB 1; Length 618;  
 Best Local Similarity 66.7%; Pred. No. 7.4;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 CDPGYIGSR 9  
 Db 455 CAPGYMGAR 463  
 RESULT 8  
 ID NC1\_HUMAN STANDARD; PRT; 2444 AA.  
 AC P46531;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1 PRECURSOR (TRANSLLOCATION-  
 DE ASSOCIATED NOTCH PROTEIN TAN-1) (FRAGMENT).  
 GN NOTCH1 OR TAN1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91347367; PubMed=1831692;  
 RA Ellisen L.W., Bird J., West D.C., Soreng A.L., Reynolds T.C.,

RA Smith S.D., Sklar J.;  
 RT "TAN-1, the human homolog of the Drosophila notch gene, is broken by  
 chromosomal translocations in T lymphoblastic neoplasms.";  
 RL Cell 66:649-661(1991).  
 CC !- FUNCTION: MAY BE IMPORTANT FOR NORMAL LYMPHOCYTE FUNCTION. IN  
 CC ALTERED FORM, MAY CONTRIBUTE TO TRANSFORMATION OR PROGRESSION  
 CC IN SOME T-CELL NEOPLASMS.  
 CC !- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC !- TISSUE SPECIFICITY: IN FETAL TISSUES MOST ABUNDANT IN SPLEEN, IT  
 CC BRAIN STEM AND LUNG. ALSO PRESENT IN MOST ADULT TISSUES WHERE IT  
 CC IS FOUND MAINLY IN LYMPHOID TISSUES.  
 CC !- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.  
 CC !- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.  
 CC !- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.  
 CC !- SIMILARITY: CONTAINS 5 ANK REPEATS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL; M73980; AAG60614.1; -.  
 DR HSP; P00740; IIXA.  
 DR MIM; 190198; -.  
 DR InterPro; IPR000152; -.  
 DR InterPro; IPR000561; -.  
 DR InterPro; IPR000800; -.  
 DR InterPro; IPR001881; -.  
 DR InterPro; IPR002110; -.  
 DR Pfam; PF00008; EGF; 36.  
 DR Pfam; PF00023; ank; 6.  
 DR Pfam; PF00066; notch; 3.  
 DR PROSITE; PS50088; ANK\_REPEAT; 4.  
 DR PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 20.  
 DR PROSITE; PS00022; EGF\_1; 34.  
 DR PROSITE; PS01186; EGF\_2; 26.  
 DR PROSITE; PS01187; EGF\_CA; 18.  
 KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;  
 KW Transmembrane; Signal; Glycoprotein.  
 FT SIGNAL 1 18 POTENTIAL.  
 FT CHAIN 19 >2444 NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 1.  
 FT DOMAIN 19 1736 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1737 1757 POTENTIAL.  
 FT DOMAIN 1758 >2444 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 20 58 EGF-LIKE 1.  
 FT DOMAIN 59 99 EGF-LIKE 2.  
 FT DOMAIN 102 139 EGF-LIKE 3.  
 FT DOMAIN 140 176 EGF-LIKE 4.  
 FT DOMAIN 178 216 EGF-LIKE 5.  
 FT DOMAIN 218 255 EGF-LIKE 6.  
 FT DOMAIN 257 293 EGF-LIKE 7.  
 FT DOMAIN 295 333 EGF-LIKE 8.  
 FT DOMAIN 335 371 EGF-LIKE 9.  
 FT DOMAIN 372 410 EGF-LIKE 10.  
 FT DOMAIN 412 450 EGF-LIKE 11.  
 FT DOMAIN 452 488 EGF-LIKE 12.  
 FT DOMAIN 490 526 EGF-LIKE 13.  
 FT DOMAIN 528 564 EGF-LIKE 14.  
 FT DOMAIN 566 601 EGF-LIKE 15.  
 FT DOMAIN 603 639 EGF-LIKE 16.  
 FT DOMAIN 641 676 EGF-LIKE 17.  
 FT DOMAIN 678 714 EGF-LIKE 18.  
 FT DOMAIN 716 751 EGF-LIKE 19.  
 FT DOMAIN 753 789 EGF-LIKE 20.  
 FT DOMAIN 791 827 EGF-LIKE 21.  
 FT DOMAIN 829 868 EGF-LIKE 22.  
 FT DOMAIN 870 906 EGF-LIKE 23.  
 FT DOMAIN 908 944 EGF-LIKE 24.  
 FT DOMAIN 946 982 EGF-LIKE 25.

FT	DOMAIN	984	1020	EGF-LIKE 26.	FT	DISULFID	612	627	BY SIMILARITY.
FT	DOMAIN	1022	1058	EGF-LIKE 27.	FT	DISULFID	629	638	BY SIMILARITY.
FT	DOMAIN	1060	1096	EGF-LIKE 28.	FT	DISULFID	645	655	BY SIMILARITY.
FT	DOMAIN	1098	1144	EGF-LIKE 29.	FT	DISULFID	650	664	BY SIMILARITY.
FT	DOMAIN	1146	1182	EGF-LIKE 30.	FT	DISULFID	666	675	BY SIMILARITY.
FT	DOMAIN	1184	1220	EGF-LIKE 31.	FT	DISULFID	682	693	BY SIMILARITY.
FT	DOMAIN	1222	1266	EGF-LIKE 32.	FT	DISULFID	687	702	BY SIMILARITY.
FT	DOMAIN	1268	1306	EGF-LIKE 33.	FT	DISULFID	704	713	BY SIMILARITY.
FT	DOMAIN	1308	1347	EGF-LIKE 34.	FT	DISULFID	720	730	BY SIMILARITY.
FT	DOMAIN	1349	1385	EGF-LIKE 35.	FT	DISULFID	741	750	BY SIMILARITY.
FT	DOMAIN	1388	1427	EGF-LIKE 36.	FT	DISULFID	757	768	BY SIMILARITY.
FT	REPEAT	1446	1481	LIN/NOTCH 1.	FT	DISULFID	777	788	BY SIMILARITY.
FT	REPEAT	1482	1523	LIN/NOTCH 2.	FT	DISULFID	795	806	BY SIMILARITY.
FT	REPEAT	1524	1563	LIN/NOTCH 3.	FT	DISULFID	800	815	BY SIMILARITY.
FT	REPEAT	1563	1600	ANK 1.	FT	DISULFID	817	826	BY SIMILARITY.
FT	REPEAT	1601	1641	ANK 2.	FT	DISULFID	833	844	BY SIMILARITY.
FT	REPEAT	1641	1681	ANK 3.	FT	DISULFID	857	867	BY SIMILARITY.
FT	REPEAT	1681	1721	ANK 4.	FT	DISULFID	874	885	BY SIMILARITY.
FT	REPEAT	1721	1761	ANK 5.	FT	DISULFID	896	905	BY SIMILARITY.
FT	DOMAIN	1761	1801	POLY-VAL.	FT	DISULFID	912	923	BY SIMILARITY.
FT	DOMAIN	1801	1841	POLY-ARG.	FT	DISULFID	932	943	BY SIMILARITY.
FT	DOMAIN	1841	1881	POLY-PRO.	FT	DISULFID	988	999	BY SIMILARITY.
FT	DOMAIN	1881	1921	POLY-ALA.	FT	DISULFID	1010	1019	BY SIMILARITY.
FT	DOMAIN	1921	1961	POLY-GLU.	FT	DISULFID	1026	1037	BY SIMILARITY.
FT	DOMAIN	1961	2001	POLY-GLN.	FT	DISULFID	1031	1046	BY SIMILARITY.
FT	DOMAIN	2001	2041	POLY-PRO.	FT	DISULFID	1048	1057	BY SIMILARITY.
FT	DOMAIN	2041	2081	BY SIMILARITY.	FT	DISULFID	1064	1075	BY SIMILARITY.
FT	DOMAIN	2081	2121	BY SIMILARITY.	FT	DISULFID	1069	1084	BY SIMILARITY.
FT	DOMAIN	2121	2161	BY SIMILARITY.	FT	DISULFID	1086	1095	BY SIMILARITY.
FT	DOMAIN	2161	2201	BY SIMILARITY.	FT	DISULFID	1102	1123	BY SIMILARITY.
FT	DOMAIN	2201	2241	BY SIMILARITY.	FT	DISULFID	1117	1132	BY SIMILARITY.
FT	DOMAIN	2241	2281	BY SIMILARITY.	FT	DISULFID	1134	1143	BY SIMILARITY.

Query Match 74.1%; Score 40; DB 1; Length 2444;

Best Local Similarity 75.0%; Pred. No. 30;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gap

QY 1 CDPGYGS 8

Db 666 CDPGYGS 673

RESULT 9

NTCL\_MOUSE

ID NTCL\_MOUSE STANDARD; PRT; 2531 AA.

AC Q01705;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR (MOTCH PROTEIN).

GN NOTCH1 OR MOTCH

OS Mus musculus (Mouse)

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;

OX NCBI\_Taxid=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Embryo;

RX MEDLINE=93194170; PubMed=8449489;

RA Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,

RA Copeland N.G., Gridley T.;

RT "Cloning, analysis, and chromosomal localization of Notch-1, a mouse

RT homolog of Drosophila Notch.";

RL Genomics 15:259-264(1993).

RN [2]

RP SEQUENCE OF 1551-2170 FROM N.A.

RC TISSUE=Embryo;



RX MEDLINE-93048835; PubMed-1425352;  
 RA Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,  
 RA Greenspan R.J., McMahon A.P., Gridley T.;  
 RT "Expression pattern of Notch, a mouse homolog of Drosophila Notch,  
 RT suggests an important role in early postimplantation mouse  
 RT development.";   
 RL Development 115:737-744(1992).  
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.  
 CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.  
 CC -!- SIMILARITY: CONTAINS 5 ANK REPEATS.  
 CC -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z11886; CAA77941.1; -;  
 DR HSP; P00740; LIXA.  
 DR MGD; MGI:97363; Notchl.  
 DR InterPro; IPR000152; -;  
 DR InterPro; IPR000561; -;  
 DR InterPro; IPR000800; -;  
 DR InterPro; IPR001438; -;  
 DR InterPro; IPR001881; -;  
 DR InterPro; IPR002110; -;  
 DR Pfam; PF00008; EGF; 35.  
 DR Pfam; PF00023; ank; 6.  
 DR Pfam; PF00066; notch; 3.  
 DR PRINTS; PS00010; EGFBLD.  
 DR PROSITE; PS00088; ANK\_REPEAT; 2.  
 DR PROSITE; PS0297; ANK\_REPEAT\_REGION; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 22.  
 DR PROSITE; PS00022; EGF\_1; 34.  
 DR PROSITE; PS01186; EGF\_2; 27.  
 DR PROSITE; PS01187; EGF\_CA; 21.  
 KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;  
 KW Transmembrane; Signal; Glycoprotein.  
 FT SIGNAL 1 18 POTENTIAL.  
 FT CHAIN 19 2531 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.  
 FT DOMAIN 19 1725 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 1726 1746 POTENTIAL.  
 FT DOMAIN 1747 2531 CYTOPLASMIC (POTENTIAL).  
 FT EGF-LIKE 1 20 58  
 FT DOMAIN 59 99 EGF-LIKE 2.  
 FT DOMAIN 102 139 EGF-LIKE 3.  
 FT DOMAIN 140 176 EGF-LIKE 4.  
 FT DOMAIN 178 216 EGF-LIKE 5.  
 FT DOMAIN 218 255 EGF-LIKE 6.  
 FT DOMAIN 257 293  
 FT DOMAIN 295 333 EGF-LIKE 7. CALCULUM-BINDING (POTENTIAL).  
 FT DOMAIN 335 371 EGF-LIKE 8. CALCULUM-BINDING (POTENTIAL).  
 FT DOMAIN 372 410 EGF-LIKE 9. CALCULUM-BINDING (POTENTIAL).  
 FT DOMAIN 412 450 EGF-LIKE 10. CALCULUM-BINDING (POTENTIAL).  
 FT DOMAIN 452 488 EGF-LIKE 11. CALCULUM-BINDING (POTENTIAL).  
 FT DOMAIN 490 526 EGF-LIKE 12. CALCULUM-BINDING (POTENTIAL).  
 FT DOMAIN 528 564 EGF-LIKE 13. CALCULUM-BINDING (POTENTIAL).  
 FT DOMAIN 566 601 EGF-LIKE 14. CALCULUM-BINDING (POTENTIAL).  
 FT DOMAIN 603 639 EGF-LIKE 15. CALCULUM-BINDING (POTENTIAL).  
 FT DOMAIN 641 676 EGF-LIKE 16. CALCULUM-BINDING (POTENTIAL).  
 FT DOMAIN 678 714 EGF-LIKE 17. CALCULUM-BINDING (POTENTIAL).  
 FT DOMAIN 716 751 EGF-LIKE 18. CALCULUM-BINDING (POTENTIAL).  
 FT DOMAIN 753 789 EGF-LIKE 19. CALCULUM-BINDING (POTENTIAL).  
 FT DOMAIN 791 827 EGF-LIKE 20. CALCULUM-BINDING (POTENTIAL).  
 FT DOMAIN 829 867 EGF-LIKE 21. CALCULUM-BINDING (POTENTIAL).  
 FT DOMAIN 869 905 EGF-LIKE 22.  
 FT DOMAIN 907 943 EGF-LIKE 23. CALCULUM-BINDING (POTENTIAL).  
 FT DOMAIN 945 EGF-LIKE 24.  
 FT EGF-LIKE 25. CALCULUM-BINDING (POTENTIAL).

FT	DOMAIN	983	1019	EGF-LIKE 26.
FT	DOMAIN	1021	1057	EGF-LIKE 27.
FT	DOMAIN	1059	1095	EGF-LIKE 28.
FT	DOMAIN	1097	1143	EGF-LIKE 29.
FT	DOMAIN	1145	1181	EGF-LIKE 30.
FT	DOMAIN	1183	1219	EGF-LIKE 31.
FT	DOMAIN	1221	1265	EGF-LIKE 32.
FT	DOMAIN	1267	1305	EGF-LIKE 33.
FT	DOMAIN	1307	1346	EGF-LIKE 34.
FT	DOMAIN	1348	1384	EGF-LIKE 35.
FT	DOMAIN	1387	1426	EGF-LIKE 36.
FT	DOMAIN	1449	1462	CYS-RICH.
FT	REPEAT	1481	1480	LIN/NOTCH 1.
FT	REPEAT	1523	1562	LIN/NOTCH 2.
FT	REPEAT	1917	1947	LIN/NOTCH 3.
FT	REPEAT	1949	1979	ANK 1.
FT	REPEAT	1983	2012	ANK 2.
FT	REPEAT	2016	2045	ANK 3.
FT	REPEAT	2049	2078	ANK 4.
FT	DISULFID	24	37	ANK 5.
FT	DISULFID	31	46	BY SIMILARITY.
FT	DISULFID	63	74	BY SIMILARITY.
FT	DISULFID	68	87	BY SIMILARITY.
FT	DISULFID	89	98	BY SIMILARITY.
FT	DISULFID	106	117	BY SIMILARITY.
FT	DISULFID	111	127	BY SIMILARITY.
FT	DISULFID	129	138	BY SIMILARITY.
FT	DISULFID	144	155	BY SIMILARITY.
FT	DISULFID	149	164	BY SIMILARITY.
FT	DISULFID	166	175	BY SIMILARITY.
FT	DISULFID	182	195	BY SIMILARITY.
FT	DISULFID	189	204	BY SIMILARITY.
FT	DISULFID	206	215	BY SIMILARITY.
FT	DISULFID	222	233	BY SIMILARITY.
FT	DISULFID	227	243	BY SIMILARITY.
FT	DISULFID	245	254	BY SIMILARITY.
FT	DISULFID	261	272	BY SIMILARITY.
FT	DISULFID	266	281	BY SIMILARITY.
FT	DISULFID	283	292	BY SIMILARITY.
FT	DISULFID	299	312	BY SIMILARITY.
FT	DISULFID	306	321	BY SIMILARITY.
FT	DISULFID	323	332	BY SIMILARITY.
FT	DISULFID	339	350	BY SIMILARITY.
FT	DISULFID	344	359	BY SIMILARITY.
FT	DISULFID	361	370	BY SIMILARITY.
FT	DISULFID	376	387	BY SIMILARITY.
FT	DISULFID	381	398	BY SIMILARITY.
FT	DISULFID	400	409	BY SIMILARITY.
FT	DISULFID	416	429	BY SIMILARITY.
FT	DISULFID	423	438	BY SIMILARITY.
FT	DISULFID	440	449	BY SIMILARITY.
FT	DISULFID	456	467	BY SIMILARITY.
FT	DISULFID	461	476	BY SIMILARITY.
FT	DISULFID	478	487	BY SIMILARITY.
FT	DISULFID	494	505	BY SIMILARITY.
FT	DISULFID	499	514	BY SIMILARITY.
FT	DISULFID	516	525	BY SIMILARITY.
FT	DISULFID	532	543	BY SIMILARITY.
FT	DISULFID	537	552	BY SIMILARITY.
FT	DISULFID	554	563	BY SIMILARITY.
FT	DISULFID	570	580	BY SIMILARITY.
FT	DISULFID	575	589	BY SIMILARITY.
FT	DISULFID	591	600	BY SIMILARITY.
FT	DISULFID	607	618	BY SIMILARITY.
FT	DISULFID	612	627	BY SIMILARITY.
FT	DISULFID	629	638	BY SIMILARITY.
FT	DISULFID	645	655	BY SIMILARITY.
FT	DISULFID	650	664	BY SIMILARITY.
FT	DISULFID	666	675	BY SIMILARITY.
FT	DISULFID	682	693	BY SIMILARITY.
FT	DISULFID	687	702	BY SIMILARITY.
FT	DISULFID	704	713	BY SIMILARITY.

FT DISULFID 720 730 BY SIMILARITY.  
 FT DISULFID 725 739 BY SIMILARITY.  
 FT DISULFID 741 750 BY SIMILARITY.  
 FT DISULFID 757 768 BY SIMILARITY.  
 FT DISULFID 762 777 BY SIMILARITY.  
 FT DISULFID 779 788 BY SIMILARITY.  
 FT DISULFID 795 806 BY SIMILARITY.  
 FT DISULFID 800 815 BY SIMILARITY.  
 FT DISULFID 817 826 BY SIMILARITY.  
 FT DISULFID 833 844 BY SIMILARITY.  
 FT DISULFID 838 855 BY SIMILARITY.  
 FT DISULFID 857 866 BY SIMILARITY.  
 FT DISULFID 873 884 BY SIMILARITY.  
 FT DISULFID 878 893 BY SIMILARITY.  
 FT DISULFID 911 922 BY SIMILARITY.  
 FT DISULFID 916 931 BY SIMILARITY.  
 FT DISULFID 933 942 BY SIMILARITY.  
 FT DISULFID 987 998 BY SIMILARITY.  
 FT DISULFID 992 1007 BY SIMILARITY.  
 FT DISULFID 1009 1018 BY SIMILARITY.  
 FT DISULFID 1025 1036 BY SIMILARITY.  
 FT DISULFID 1030 1045 BY SIMILARITY.  
 FT DISULFID 1047 1056 BY SIMILARITY.  
 FT DISULFID 1063 1074 BY SIMILARITY.  
 FT DISULFID 1068 1083 BY SIMILARITY.  
 FT DISULFID 1085 1094 BY SIMILARITY.  
 FT DISULFID 1101 1122 BY SIMILARITY.  
 FT DISULFID 1116 1131 BY SIMILARITY.  
 FT DISULFID 1133 1142 BY SIMILARITY.  
 FT DISULFID 1149 1160 BY SIMILARITY.

Query Match 74.1%; Score 40; DB 1; Length 2531;  
 Best Local Similarity 75.0%; Pred. No. 31;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDPGYIGS 8  
 : : : : :  
 Db 666 CEPGYIGS 673

RESULT 10  
 NTC1\_RAT  
 ID NTC1\_RAT STANDARD; PRT: 2531 AA.  
 AC Q07008;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR.  
 GN NOTCH1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_Taxid=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Schwann cell;  
 RX MEDLINE=92111383; PubMed=1764995;  
 RA Weinmaster G., Roberts V.J., Lemke G.;  
 RT "A homolog of Drosophila Notch expressed during mammalian  
 development.";  
 RL Development 113:199-205(1991).  
 RN [2]  
 RP REVISIONS TO 1652-1653.  
 RA Weinmaster G.;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: REQUIRED FOR THE CORRECT DIFFERENTIATION OF A NUMBER  
 OF TISSUES.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- DEVELOPMENTAL STAGE: IN THE EMBRYO, HIGHEST LEVELS OCCUR BETWEEN  
 DAYS 12 AND 14 AND DECREASE RAPIDLY TO MUCH LOWER LEVELS IN THE  
 ADULT.  
 CC -1- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.

CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 3 LINK/NOTCH REPEATS.  
 CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL; X57405; CAA40667.1; -  
 DR HSP: P00740; 11XA.  
 DR InterPro; IPR000152;  
 DR InterPro; IPR000561;  
 DR InterPro; IPR000800;  
 DR InterPro; IPR001438;  
 DR InterPro; IPR001881;  
 DR InterPro; IPR002049;  
 DR InterPro; IPR002110;  
 DR Pfam; PF00008; EGF; 36.  
 DR Pfam; PF00023; ank; 6.  
 DR Pfam; PF00066; notch; 3.  
 DR PRINTS; PR00010; EGFBL00D.  
 DR PRINTS; PR00011; EGFLAMIN.  
 DR PROSITE; PS00088; ANK\_REPEAT; 4.  
 DR PROSITE; PS0297; ANK\_REPEAT\_REGION; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 22.  
 DR PROSITE; PS00022; EGF\_1; 35.  
 DR PROSITE; PS01186; EGF\_2; 26.  
 DR PROSITE; PS01187; EGF\_CA; 21.  
 KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;  
 KW Transmembrane; Signal; Glycoprotein.  
 FT SIGNAL 1 18  
 FT CHAIN 19 2531  
 FT DOMAIN 19 1723  
 FT TRANSMEM 1724 1746  
 FT DOMAIN 1747 2531  
 FT DOMAIN 20 58  
 FT DOMAIN 59 99  
 FT DOMAIN 102 139  
 FT DOMAIN 140 176  
 FT DOMAIN 178 216  
 FT DOMAIN 218 255  
 FT DOMAIN 257 293  
 FT DOMAIN 295 333  
 FT DOMAIN 335 371  
 FT DOMAIN 372 410  
 FT DOMAIN 412 450  
 FT DOMAIN 452 488  
 FT DOMAIN 490 526  
 FT DOMAIN 528 564  
 FT DOMAIN 566 601  
 FT DOMAIN 603 639  
 FT DOMAIN 641 676  
 FT DOMAIN 678 714  
 FT DOMAIN 716 751  
 FT DOMAIN 753 789  
 FT DOMAIN 791 827  
 FT DOMAIN 829 867  
 FT DOMAIN 869 905  
 FT DOMAIN 907 943  
 FT DOMAIN 945 981  
 FT DOMAIN 983 1019  
 FT DOMAIN 1021 1057  
 FT DOMAIN 1059 1095  
 FT DOMAIN 1097 1143  
 FT DOMAIN 1145 1181  
 FT DOMAIN 1183 1219  
 FT DOMAIN 1221 1265  
 FT DOMAIN 1267 1305  
 FT DOMAIN 1307 1346

NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.  
 EXTRACELLULAR (POTENTIAL).  
 POTENTIAL.  
 CYTOPLASMIC (POTENTIAL).  
 EGF-LIKE 1.  
 EGF-LIKE 2.  
 EGF-LIKE 3.  
 EGF-LIKE 4.  
 EGF-LIKE 5.  
 EGF-LIKE 6.  
 EGF-LIKE 7.  
 EGF-LIKE 8.  
 EGF-LIKE 9.  
 EGF-LIKE 10.  
 EGF-LIKE 11.  
 EGF-LIKE 12.  
 EGF-LIKE 13.  
 EGF-LIKE 14.  
 EGF-LIKE 15.  
 EGF-LIKE 16.  
 EGF-LIKE 17.  
 EGF-LIKE 18.  
 EGF-LIKE 19.  
 EGF-LIKE 20.  
 EGF-LIKE 21.  
 EGF-LIKE 22.  
 EGF-LIKE 23.  
 EGF-LIKE 24.  
 EGF-LIKE 25.  
 EGF-LIKE 26.  
 EGF-LIKE 27.  
 EGF-LIKE 28.  
 EGF-LIKE 29.  
 EGF-LIKE 30.  
 EGF-LIKE 31.  
 EGF-LIKE 32.  
 EGF-LIKE 33.  
 EGF-LIKE 34.

FT DOMAIN 1348 EGF-LIKE 35.  
 FT DOMAIN 1349 EGF-LIKE 36.  
 FT REPEAT 1449 CYS-RICH.  
 FT REPEAT 1917 ANK 1.  
 FT REPEAT 1950 ANK 2.  
 FT REPEAT 1984 ANK 3.  
 FT REPEAT 2017 ANK 4.  
 FT REPEAT 2050 ANK 5.  
 FT DISULFID 24 37 BY SIMILARITY.  
 FT DISULFID 31 46 BY SIMILARITY.  
 FT DISULFID 48 57 BY SIMILARITY.  
 FT DISULFID 63 74 BY SIMILARITY.  
 FT DISULFID 68 87 BY SIMILARITY.  
 FT DISULFID 89 98 BY SIMILARITY.  
 FT DISULFID 106 117 BY SIMILARITY.  
 FT DISULFID 111 127 BY SIMILARITY.  
 FT DISULFID 129 138 BY SIMILARITY.  
 FT DISULFID 144 155 BY SIMILARITY.  
 FT DISULFID 149 164 BY SIMILARITY.  
 FT DISULFID 166 175 BY SIMILARITY.  
 FT DISULFID 182 195 BY SIMILARITY.  
 FT DISULFID 189 204 BY SIMILARITY.  
 FT DISULFID 206 215 BY SIMILARITY.  
 FT DISULFID 222 233 BY SIMILARITY.  
 FT DISULFID 227 243 BY SIMILARITY.  
 FT DISULFID 245 254 BY SIMILARITY.  
 FT DISULFID 261 272 BY SIMILARITY.  
 FT DISULFID 266 281 BY SIMILARITY.  
 FT DISULFID 283 292 BY SIMILARITY.  
 FT DISULFID 299 312 BY SIMILARITY.  
 FT DISULFID 306 321 BY SIMILARITY.  
 FT DISULFID 323 332 BY SIMILARITY.  
 FT DISULFID 339 350 BY SIMILARITY.  
 FT DISULFID 344 359 BY SIMILARITY.  
 FT DISULFID 361 370 BY SIMILARITY.  
 FT DISULFID 376 387 BY SIMILARITY.  
 FT DISULFID 381 398 BY SIMILARITY.  
 FT DISULFID 400 409 BY SIMILARITY.  
 FT DISULFID 416 429 BY SIMILARITY.  
 FT DISULFID 423 438 BY SIMILARITY.  
 FT DISULFID 440 449 BY SIMILARITY.  
 FT DISULFID 456 467 BY SIMILARITY.  
 FT DISULFID 461 476 BY SIMILARITY.  
 FT DISULFID 478 487 BY SIMILARITY.  
 FT DISULFID 494 505 BY SIMILARITY.  
 FT DISULFID 499 514 BY SIMILARITY.  
 FT DISULFID 516 525 BY SIMILARITY.  
 FT DISULFID 532 543 BY SIMILARITY.  
 FT DISULFID 537 552 BY SIMILARITY.  
 FT DISULFID 554 563 BY SIMILARITY.  
 FT DISULFID 570 580 BY SIMILARITY.  
 FT DISULFID 575 589 BY SIMILARITY.  
 FT DISULFID 591 600 BY SIMILARITY.  
 FT DISULFID 607 618 BY SIMILARITY.  
 FT DISULFID 612 627 BY SIMILARITY.  
 FT DISULFID 629 638 BY SIMILARITY.  
 FT DISULFID 645 655 BY SIMILARITY.  
 FT DISULFID 650 664 BY SIMILARITY.  
 FT DISULFID 666 675 BY SIMILARITY.  
 FT DISULFID 682 693 BY SIMILARITY.  
 FT DISULFID 687 702 BY SIMILARITY.  
 FT DISULFID 704 713 BY SIMILARITY.  
 FT DISULFID 720 730 BY SIMILARITY.  
 FT DISULFID 725 739 BY SIMILARITY.  
 FT DISULFID 741 750 BY SIMILARITY.  
 FT DISULFID 757 768 BY SIMILARITY.  
 FT DISULFID 762 777 BY SIMILARITY.  
 FT DISULFID 779 788 BY SIMILARITY.  
 FT DISULFID 795 806 BY SIMILARITY.  
 FT DISULFID 800 815 BY SIMILARITY.  
 FT DISULFID 817 826 BY SIMILARITY.  
 FT DISULFID 833 844 BY SIMILARITY.  
 FT DISULFID 838 855 BY SIMILARITY.

FT DISULFID 857 866 BY SIMILARITY.  
 FT DISULFID 873 884 BY SIMILARITY.  
 FT DISULFID 878 893 BY SIMILARITY.  
 FT DISULFID 895 904 BY SIMILARITY.  
 FT DISULFID 911 922 BY SIMILARITY.  
 FT DISULFID 916 931 BY SIMILARITY.  
 FT DISULFID 933 942 BY SIMILARITY.  
 FT DISULFID 987 998 BY SIMILARITY.  
 FT DISULFID 992 1007 BY SIMILARITY.  
 FT DISULFID 1009 1018 BY SIMILARITY.  
 FT DISULFID 1025 1036 BY SIMILARITY.  
 FT DISULFID 1030 1045 BY SIMILARITY.  
 FT DISULFID 1047 1056 BY SIMILARITY.  
 FT DISULFID 1063 1074 BY SIMILARITY.  
 FT DISULFID 1068 1083 BY SIMILARITY.  
 FT DISULFID 1085 1094 BY SIMILARITY.  
 FT DISULFID 1101 1122 BY SIMILARITY.  
 FT DISULFID 1116 1131 BY SIMILARITY.  
 FT DISULFID 1133 1142 BY SIMILARITY.  
 FT DISULFID 1149 1160 BY SIMILARITY.  
 FT DISULFID 1154 1169 BY SIMILARITY.  
 FT DISULFID 1171 1180 BY SIMILARITY.  
 FT DISULFID 1187 1198 BY SIMILARITY.  
 FT DISULFID 1192 1207 BY SIMILARITY.

Query Match 74.1%; Score 40; DB 1; Length 2531;  
 Best Local Similarity 75.0%; Pred. No. 31;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDPGYIGS 8  
 I:|||| I|  
 Db 666 CEPGYTGS 673

## RESULT 11

ID NOTC\_DROME STANDARD; PRT: 2703 AA.  
 AC P07207; P04154;  
 DT 01-NOV-1986 (Rel. 03, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE NEUROGENIC LOCUS NOTCH PROTEIN PRECURSOR.  
 GN N.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86079539; PubMed=3935325;  
 RA Wharton K.A., Johansen K.M., Xu T., Artavanis-Tsakonas S.;  
 "Nucleotide sequence from the neurogenic locus notch implies a gene product that shares homology with proteins containing EGF-like repeats.";  
 RL Cell 43:567-581(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OREGON-R;  
 RX MEDLINE=87064624; PubMed=3097517;  
 RA Kidd S., Kelley M.R., Young M.W.;  
 "Sequence of the notch locus of Drosophila melanogaster: relationship of the encoded protein to mammalian clotting and growth factors.";  
 RL Mol. Cell. Biol. 6:3094-3108(1986).  
 RN [3]  
 RP SEQUENCE OF 2505-2611 FROM N.A.  
 RX MEDLINE=85099329; PubMed=2981631;  
 RA Wharton K.A., Yedvobnick B., Finnerty V.G., Artavanis-Tsakonas S.;  
 "opa: a novel family of transcribed repeats shared by the Notch locus and other developmentally regulated loci in D. melanogaster.";  
 RL Cell 40:55-62(1985).  
 RN [4]  
 RP SEQUENCE OF 1-8 FROM N.A.



FT DISULFID 476 485 BY SIMILARITY.  
 FT DISULFID 492 503 BY SIMILARITY.  
 FT DISULFID 512 512 BY SIMILARITY.  
 FT DISULFID 514 523 BY SIMILARITY.  
 FT DISULFID 530 541 BY SIMILARITY.  
 FT DISULFID 535 550 BY SIMILARITY.  
 FT DISULFID 552 561 BY SIMILARITY.  
 FT DISULFID 568 579 BY SIMILARITY.  
 FT DISULFID 573 588 BY SIMILARITY.  
 FT DISULFID 590 599 BY SIMILARITY.  
 FT DISULFID 606 616 BY SIMILARITY.  
 FT DISULFID 611 625 BY SIMILARITY.  
 FT DISULFID 627 636 BY SIMILARITY.  
 FT DISULFID 643 654 BY SIMILARITY.  
 FT DISULFID 648 663 BY SIMILARITY.  
 FT DISULFID 665 674 BY SIMILARITY.  
 FT DISULFID 681 692 BY SIMILARITY.

Query Match 74.18; Score 40; DB 1; Length 2703;

Best Local Similarity 85.78; Pred. No. 33;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 CDPGYIG 7

Db 627 CDPGYIG 633

# RESULT 12

## TENX\_HUMAN

ID TENX\_HUMAN STANDARD; PRT: 4289 AA.  
 AC P22105; P78530; P78531; Q08424; Q9UMG7;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE TENASCIN-X PRECURSOR (TN-X) (HEXABRACHION-LIKE).  
 GN TENX OR TNX OR XB OR HXBL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Rowen L., Dinkers C., Baskin D., Faust J., Loretz C., Ahearn M.E.,  
 RA Banta A., Schwartz S., Smith T.M., Spies T., Hood L.;  
 RT "Sequence determination of 300 kilobases of the human class III  
 RT MHC locus."  
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]

SEQUENCE OF 1-747 AND 1687-1944 FROM N.A.

RC TISSUE=Leukocyte;  
 RX MEDLINE=93300909; PubMed=7686164;  
 RA Bristow J., Tee M.K., Gitelman S.E., Mellon S.H., Miller W.L.;  
 RT "Tenascin-X: a novel extracellular matrix protein encoded by the human  
 RT XB gene overlapping P450c21B."  
 RL J. Cell Biol. 122:265-278(1993).  
 RN [3]

SEQUENCE FROM N.A. (ISOFORM XB-SHORT).  
 RC TISSUE=Adrenal gland;  
 RX MEDLINE=96015044; PubMed=8530023;

RA Tee M.K., Thomson A.A., Bristow J., Miller W.L.;  
 RT "Sequences promoting the transcription of the human XA gene  
 RT overlapping P450c21A correctly predict the presence of a novel,  
 RT adrenal-specific, truncated form of tenascin-X";  
 RL Genomics 28:171-178(1995).  
 RN [4]

SEQUENCE OF 1-23 FROM N.A.  
 RC TISSUE=Fetal adrenal gland;  
 RX MEDLINE=97081760; PubMed=8923003;

RA Speck M., Barry F., Miller W.L.;  
 RT "Alternate promoters and alternate splicing of human tenascin-X, a  
 RT gene with 5' and 3' ends buried in other genes."  
 RL Hum. Mol. Genet. 5:1749-1758(1996).  
 RN [5]

RP

SEQUENCE OF 3470-4289 FROM N.A.

RX MEDLINE=89367293; PubMed=2475872;  
 RA Morel Y., Bristow J., Gitelman S.E., Miller W.L.;  
 RT "transcript encoded on the opposite strand of the human steroid 21-  
 RT hydroxylase/complement component C4 gene locus";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:6582-6586(1989).  
 CC -!- FUNCTION: APPEARS TO MEDIATE INTERACTIONS BETWEEN CELLS AND THE  
 CC EXTRACELLULAR MATRIX. SUBSTRATE-ADHESION MOLECULE THAT APPEARS TO  
 CC INHIBIT CELL MIGRATION. MAY PLAY A ROLE IN SUPPORTING THE GROWTH  
 CC OF EPITHELIAL TUMORS.  
 CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.  
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS: XB (SHOWN HERE) AND  
 CC XB-SHORT; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN FETAL ADRENAL, IN FETAL  
 CC TESTIS, FETAL SMOOTH, STRIATED AND CARDIAC MUSCLE. XB-SHORT IS  
 CC ONLY EXPRESSED IN THE ADRENAL GLAND.  
 CC -!- DISEASE: ASSOCIATION WITH CONGENITAL ADRENAL HYPERPLASIA AND EHLER  
 CC AND DANLO'S SYNDROME.  
 CC -!- SIMILARITY: CONTAINS 19 EGF-LIKE DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 32 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC -!- SIMILARITY: CONTAINS A DOMAIN RELATED TO THE C-TERMINAL PART OF  
 CC THE BETA AND GAMMA CHAINS OF FIBRINOGEN.  
 CC -!- CAUTION: THERE ARE TWO GENES FOR TN-X: TNXA AND TNXB. TNXA IS  
 CC A PARTIAL GENE WHICH CAN SOMETIME RECOMBINE WITH TNXB.

CC

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

CC

EMBL: U89337; AAB47488.1; -

DR

EMBL: AF019413; AAB67981.1; -

DR

EMBL: X71923; CAAS0739.1; -

DR

EMBL: Y13782; CAA74109.1; -

DR

EMBL: Y13783; CAA74110.1; -

DR

EMBL: U24488; AAB41287.1; -

DR

EMBL: U52696; AAC50889.1; -

DR

EMBL: M25813; AAA35884.1; -

DR

MIM: 600261; -

DR

MIM: 600985; -

DR

HSSP: P02671; 1FZD.

DR

InterPro: IPR000561; -

DR

InterPro: IPR001777; -

DR

InterPro: IPR002181; -

DR

Pfam: PF00008; EGF\_17.

DR

Pfam: PF00041; fn3; 32.

DR

Pfam: PF00147; fibrinogen\_C; 1.

DR

PROSITE: PS00022; EGF\_1; 18.

DR

PROSITE: PS01186; EGF\_2; 19.

DR

PROSITE: PS00514; FIBRINAG\_C\_DOMAIN; 1.

DR

Glycoprotein; Cell adhesion; Repeat; EGF-like domain; Coiled coil;  
 KW Extracellular matrix; Alternative splicing; Signal.

KW

Extracellular matrix; Alternative splicing; Signal.

FT

SIGNAL 1 23 POTENTIAL.

FT

CHAIN 24 4289 TENASCIN-X.

FT

DOMAIN 156 167 EGF-LIKE 1.

FT

DOMAIN 202 213 EGF-LIKE 2.

FT

DOMAIN 233 244 EGF-LIKE 3.

FT

DOMAIN 264 275 EGF-LIKE 4.

FT

DOMAIN 295 306 EGF-LIKE 5.

FT

DOMAIN 326 337 EGF-LIKE 6.

FT

DOMAIN 357 368 EGF-LIKE 7.

FT

DOMAIN 386 399 EGF-LIKE 8.

FT

DOMAIN 419 430 EGF-LIKE 9.

FT

DOMAIN 450 461 EGF-LIKE 10.

FT

DOMAIN 481 492 EGF-LIKE 11.

FT

DOMAIN 512 523 EGF-LIKE 12.

FT

DOMAIN 543 554 EGF-LIKE 13.

FT

DOMAIN 574 585 EGF-LIKE 14.

FT

DOMAIN 605 616 EGF-LIKE 15.

FT

DOMAIN 636 647 EGF-LIKE 16.



```
FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 224 224 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 294 294 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 511 511 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 562 562 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 609 609 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 708 708 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 810 AA; 89212 MW; 46F09C466AFAB0B CRC64;

Query Match 72.2%; Score 39; DB 1; Length 810;
Best Local Similarity 62.5%; Pred. No. 15;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 CDPGYIGS 8
    | |||:|
Db 504 CQPGVGN 511

RESULT 14
GLP1_CAEEL STANDARD; PRT; 1295 AA.
AC P13508;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE GLP-1 PROTEIN PRECURSOR.
GN GLP-1 OR EMB-33 OR F02A9.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fraser A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkhan R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Watson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RX "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RX elegans.";
RL Nature 368:32-38(1994).
RN [3]
RP DELETION OF 1174-1295.
RX MEDLINE=91351288; PubMed=1881436;
RA Mango S.E., Maine E.M., Kimble J.;
RA "Carboxy-terminal truncation activates glp-1 protein to specify
RA vulval fates in Caenorhabditis elegans.";
RL Nature 352:811-815(1991).
RN [4]
RP CHARACTERIZATION OF FUNCTION OF THE ANK-REPEATS.
RX MEDLINE=93354444; PubMed=8350921;
RA Roehl H., Kimble J.;
RA "Control of cell fate in C. elegans by a GLP-1 peptide consisting
RA primarily of ankyrin repeats.";
RL
```

```
RL Nature 364:632-635(1993).
RN [5]
RP FUNCTION.
RX MEDLINE=94208066; PubMed=8156602;
RA Mello C.C., Draper B.W., Priess J.R.;
RT "The maternal genes apx-1 and glp-1 and establishment of
RT dorsal-ventral polarity in the early C. elegans embryo.";
RL Cell 77:95-106(1994).
CC -!- FUNCTION: INVOLVED IN THE SPECIFICATION OF THE CELL FATES OF THE
CC BLASTOMERES, ABA AND APA. PROPER SIGNALING BY GLP-1 INDUCES ABA
CC DESCENDANTS TO PRODUCE ANTERIOR PHARYNGEAL CELLS, AND APA
CC DESCENDANTS TO ADOPT A DIFFERENT FATE. CONTRIBUTES TO THE
CC ESTABLISHMENT THE DORSAL-VENTRAL AXIS IN EARLY EMBRYOS.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- DEVELOPMENTAL STAGE: ACTS ON ABP DEVELOPMENT DURING 4-CELL AND
CC 12-CELL STAGES, AND ON ABA DEVELOPMENT DURING 12-CELL AND 28-CELL
CC STAGES.
CC -!- SIMILARITY: HIGH, TO C.ELEGANS LIN-12.
CC -!- SIMILARITY: CONTAINS 10 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -!- SIMILARITY: CONTAINS 5 ANK REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M25580; AAA28058.1; -.
CC EMBL; Z19555; CAA79620.1; -.
CC EMBL; Z29116; CAA79620.1; JOINED.
CC EMBL; Z29116; CAA82373.1; -.
CC EMBL; Z19555; CAA82373.1; JOINED.
CC PIR; A32901; A32901.
CC HSP; P00740; IIXA.
CC WormPep; F02A9.6; CE00237.
CC InterPro; IPR000152; -.
CC InterPro; IPR000561; -.
CC InterPro; IPR000800; -.
CC InterPro; IPR001881; -.
CC InterPro; IPR002110; -.
CC Pfam; PF00008; EGF; 10.
CC Pfam; PF00023; ank; 4.
CC Pfam; PF00066; notch; 3.
CC PROSITE; PS50088; ANK_REPEAT; 3.
CC PROSITE; PS50297; ANK_REPEAT_REGION; 1.
CC PROSITE; PS00010; ASX_HYDROXYL; 2.
CC PROSITE; PS00022; EGF_1; 10.
CC PROSITE; PS01186; EGF_2; 8.
CC PROSITE; PS01187; EGF_CA; 1.
CC Differentiation; Repeat; ANK repeat; EGF-like domain; Transmembrane;
KW Glycoprotein; Signal.
FT SIGNAL 1 15
FT CHAIN 16 1295
FT DOMAIN 16 764
FT TRANSMEM 765 786
FT DOMAIN 787 1295
FT DOMAIN 19 58
FT DOMAIN 117 152
FT DOMAIN 154 190
FT DOMAIN 190 230
FT DOMAIN 232 269
FT DOMAIN 271 308
FT DOMAIN 316 359
FT DOMAIN 369 406
FT DOMAIN 407 443
FT DOMAIN 446 479
FT DOMAIN 493 527
FT REPEAT 528 568
FT REPEAT 569 608
FT REPEAT 609 990
FT REPEAT 991 1.
```





FT	DOMAIN	764	800	EGF-LIKE 20.
FT	DOMAIN	803	839	EGF-LIKE 21.
FT	DOMAIN	841	877	EGF-LIKE 22.
FT	DOMAIN	878	924	EGF-LIKE 23.
FT	DOMAIN	926	962	EGF-LIKE 24.
FT	DOMAIN	964	1000	EGF-LIKE 25.
FT	DOMAIN	1002	1040	EGF-LIKE 26.
FT	DOMAIN	1042	1081	EGF-LIKE 27.
FT	DOMAIN	1083	1122	EGF-LIKE 28.
FT	DOMAIN	1126	1167	EGF-LIKE 29.
FT	REPEAT	1168	1208	LIN/NOTCH 1.
FT	REPEAT	1209	1242	LIN/NOTCH 2.
FT	REPEAT	1243	1282	LIN/NOTCH 3.
FT	REPEAT	1628	1657	ANK 1.
FT	REPEAT	1661	1691	ANK 2.
FT	REPEAT	1695	1724	ANK 3.
FT	REPEAT	1728	1757	ANK 4.
FT	REPEAT	1761	1790	ANK 5.
FT	DISULFID	25	38	BY SIMILARITY.
FT	DISULFID	32	48	BY SIMILARITY.
FT	DISULFID	50	59	BY SIMILARITY.
FT	DISULFID	65	77	BY SIMILARITY.
FT	DISULFID	71	100	BY SIMILARITY.
FT	DISULFID	102	111	BY SIMILARITY.
FT	DISULFID	119	130	BY SIMILARITY.
FT	DISULFID	124	140	BY SIMILARITY.
FT	DISULFID	142	151	BY SIMILARITY.
FT	DISULFID	157	168	BY SIMILARITY.
FT	DISULFID	162	177	BY SIMILARITY.
FT	DISULFID	179	188	BY SIMILARITY.
FT	DISULFID	195	208	BY SIMILARITY.
FT	DISULFID	202	217	BY SIMILARITY.
FT	DISULFID	219	228	BY SIMILARITY.
FT	DISULFID	235	246	BY SIMILARITY.
FT	DISULFID	240	259	BY SIMILARITY.
FT	DISULFID	261	270	BY SIMILARITY.
FT	DISULFID	277	288	BY SIMILARITY.
FT	DISULFID	282	297	BY SIMILARITY.
FT	DISULFID	299	308	BY SIMILARITY.
FT	DISULFID	315	329	BY SIMILARITY.
FT	DISULFID	323	338	BY SIMILARITY.
FT	DISULFID	340	349	BY SIMILARITY.
FT	DISULFID	356	367	BY SIMILARITY.
FT	DISULFID	361	376	BY SIMILARITY.
FT	DISULFID	378	387	BY SIMILARITY.
FT	DISULFID	393	404	BY SIMILARITY.
FT	DISULFID	398	415	BY SIMILARITY.
FT	DISULFID	417	426	BY SIMILARITY.
FT	DISULFID	433	449	BY SIMILARITY.
FT	DISULFID	443	458	BY SIMILARITY.
FT	DISULFID	460	469	BY SIMILARITY.
FT	DISULFID	476	487	BY SIMILARITY.
FT	DISULFID	481	496	BY SIMILARITY.
FT	DISULFID	498	507	BY SIMILARITY.
FT	DISULFID	514	525	BY SIMILARITY.
FT	DISULFID	519	534	BY SIMILARITY.
FT	DISULFID	536	545	BY SIMILARITY.
FT	DISULFID	552	563	BY SIMILARITY.
FT	DISULFID	557	572	BY SIMILARITY.
FT	DISULFID	574	583	BY SIMILARITY.
FT	DISULFID	590	601	BY SIMILARITY.
FT	DISULFID	595	610	BY SIMILARITY.
FT	DISULFID	612	621	BY SIMILARITY.
FT	DISULFID	626	637	BY SIMILARITY.
FT	DISULFID	631	646	BY SIMILARITY.
FT	DISULFID	648	655	BY SIMILARITY.
FT	DISULFID	662	669	BY SIMILARITY.
FT	DISULFID	664	674	BY SIMILARITY.
FT	DISULFID	676	685	BY SIMILARITY.
FT	DISULFID	692	703	BY SIMILARITY.
FT	DISULFID	697	712	BY SIMILARITY.
FT	DISULFID	714	723	BY SIMILARITY.
FT	DISULFID	730	741	BY SIMILARITY.

FT	DISULFID	735	750	BY SIMILARITY.
FT	DISULFID	752	761	BY SIMILARITY.
FT	DISULFID	768	779	BY SIMILARITY.
FT	DISULFID	773	788	BY SIMILARITY.
FT	DISULFID	790	799	BY SIMILARITY.
FT	DISULFID	807	818	BY SIMILARITY.
FT	DISULFID	812	827	BY SIMILARITY.
FT	DISULFID	829	838	BY SIMILARITY.
FT	DISULFID	845	856	BY SIMILARITY.
FT	DISULFID	850	865	BY SIMILARITY.
FT	DISULFID	867	876	BY SIMILARITY.
FT	DISULFID	882	903	BY SIMILARITY.
FT	DISULFID	897	912	BY SIMILARITY.
FT	DISULFID	914	923	BY SIMILARITY.
FT	DISULFID	930	941	BY SIMILARITY.
FT	DISULFID	935	950	BY SIMILARITY.
FT	DISULFID	952	961	BY SIMILARITY.
FT	DISULFID	968	979	BY SIMILARITY.
FT	DISULFID	973	988	BY SIMILARITY.
FT	DISULFID	990	999	BY SIMILARITY.
FT	DISULFID	1006	1019	BY SIMILARITY.
FT	DISULFID	1011	1028	BY SIMILARITY.
FT	DISULFID	1030	1039	BY SIMILARITY.
FT	DISULFID	1046	1057	BY SIMILARITY.
FT	DISULFID	1051	1069	BY SIMILARITY.
FT	DISULFID	1071	1080	BY SIMILARITY.
FT	DISULFID	1087	1098	BY SIMILARITY.
FT	DISULFID	1092	1110	BY SIMILARITY.
FT	DISULFID	1112	1121	BY SIMILARITY.
FT	DISULFID	1130	1142	BY SIMILARITY.
FT	DISULFID	1136	1155	BY SIMILARITY.
FT	DISULFID	1157	1166	BY SIMILARITY.

Query Match 72.2%; Score 39; DB 1; Length 1964;  
Best Local Similarity 77.8%; Pred. No. 36;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9  
I I I I I I I I  
Db 460 CLPGYIGSR 468

Search completed: August 15, 2001, 10:55:55  
Job time: 267 sec

Wed Aug 15 10:57:50 2001

us-09-673-785a-4.rsp

Page 18

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 15, 2001, 10:55:35 ; Search time 33.36 Seconds  
(without alignments)  
35.694 Million cell updates/sec

Title: US-09-673-785A-4  
Perfect score: 54  
Sequence: 1 CDPGYIGSR 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL16:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_unclassified:\*  
13: sp\_vertebrate:\*  
14: sp\_virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	49	90.7	816	11	070309
2	44	81.5	791	6	Q9GK49
3	42	77.8	446	5	Q9Y0A4
4	42	77.8	502	4	Q9H3S8
5	42	77.8	830	4	O43701
6	42	77.8	830	4	Q14162
7	42	77.8	955	4	Q99466
8	42	77.8	1827	5	Q20535
9	42	77.8	1999	4	Q99940
10	42	77.8	2003	4	O00306
11	42	77.8	2653	5	Q25253
12	41	75.9	372	5	Q21756
13	41	75.9	838	5	Q9VQA9
14	41	75.9	1664	5	Q9TVQ2
15	41	75.9	2524	5	Q9GPA5
16	40	74.1	372	11	O63762
17	40	74.1	374	5	Q9VPJ0
18	40	74.1	510	4	Q9H4V1
19	40	74.1	783	5	Q9Y0F8

20	40	74.1	799	5	Q27591
21	40	74.1	799	5	Q9VIG7
22	40	74.1	877	5	Q9U2D5
23	40	74.1	2144	11	Q9QYP2
24	40	74.1	2408	4	Q92566
25	40	74.1	2703	5	Q9W4T8
26	40	74.1	2704	5	Q97458
27	40	74.1	2923	4	Q9HCU4
28	40	74.1	4135	6	O18977
29	40	74.1	4288	4	O9NPK9
30	39	72.2	52	5	O9TX97
31	39	72.2	79	2	O9PD22
32	39	72.2	159	5	Q9VWNO
33	39	72.2	220	11	Q63404
34	39	72.2	615	13	O93575
35	39	72.2	664	13	Q9IAT6
36	39	72.2	1574	11	O88281
37	39	72.2	1964	11	O35442
38	39	72.2	2126	4	O9UDQ2
39	39	72.2	3209	13	O93574
40	39	72.2	3460	4	P78509
41	39	72.2	3461	11	Q60841
42	39	72.2	4006	11	O35452
43	39	72.2	4114	11	O54796
44	38	70.4	56	6	O18962
45	38	70.4	178	6	Q9TTC5

#### ALIGNMENTS

RESULT 1

ID 070309 PRELIMINARY; PRT; 816 AA.  
AC 070309: 070308: O88347;  
DT 01-AUG-1998 (TREMBLrel. 07, Created)  
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE INTEGRIN BETA-5 PRECURSOR (BETA5B INTEGRIN).  
GN ITGB5.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND TISSUE SPECIFICITY.  
RC TISSUE=LIVER;  
RX MEDLINE=98198405; PubMed=9531507;  
RA Zhang H., Tan S.M., Lu J.;  
RT "cDNA cloning reveals two mouse beta5 integrin transcripts distinct in  
RT cytoplasmic domains as a result of alternative splicing.";  
RL Biochem. J. 331:631-637(1998).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM A).  
RC TISSUE=BRAIN;  
RX MEDLINE=99098874; PubMed=9880508;  
RA Feng X., Teitelbaum S.L., Quiroz M.E., Towler D.A., Ross F.P.;  
RT "Cloning of the murine beta5 integrin subunit promoter. Identification  
RT of a novel sequence mediating granulocyte-macrophage colony-  
RT stimulating factor-dependent repression of beta5 integrin gene  
transcription.";  
RL J. Biol. Chem. 274:1366-1374(1999).  
CC -!- FUNCTION: INTEGRINS ARE A LARGE FAMILY OF CELL SURFACE  
CC GLYCOPROTEINS THAT MEDIATE CELL TO CELL & CELL TO MATRIX ADHESION.  
CC -!- SUBUNIT: DIMER OF AN ALPHA AND BETA SUBUNIT. BETA-5 ASSOCIATES  
CC WITH ALPHA-V.  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; B (SHOWN HERE) AND A; ARE  
CC PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- TISSUE SPECIFICITY: BOTH ISOFORMS ARE EXPRESSED IN ALL TISSUES  
CC EXAMINED: LIVER, BRAIN, COLON, KIDNEY, LONG, SPLEEN, UTERUS,  
CC SKELETAL MUSCLE AND HEART. VERY LOW LEVELS OF BOTH ISOFORMS ARE  
CC FOUND IN SPLEEN. LOW LEVELS OF ISOFORM B ARE ALSO FOUND IN UTERUS

CC AND SKELETAL MUSCLE. IN ALL TISSUES TESTED, ISOFORM B IS DETECTED  
 CC AT MUCH LOWER LEVELS THAN ISOFORM A.  
 CC -1- PTM: THE CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN DISULFIDE  
 CC BONDS.  
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.  
 DR EMBL: AF043257; AAC40110.1; -;  
 DR EMBL: AF043256; AAC40109.1; -;  
 DR EMBL: AF022110; AAD08782.1; -;  
 DR MGI: 96614; Itgbs.  
 DR InterPro: IPR000561; -;  
 DR InterPro: IPR001169; -;  
 DR InterPro: IPR002369; -;  
 DR Pfam: PF00362; Integrin\_B; 1.  
 DR ProDom: PD001611; -; 1.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_2.  
 DR PROSITE: PS01186; EGF\_2; 2.  
 DR PROSITE: PS00243; INTEGRIN\_BETA; 2.  
 KW Integrin; Cell adhesion; Transmembrane; Glycoprotein; Repeat;  
 FT Extracellular matrix; Cytoskeleton; Signal; Alternative splicing.  
 FT SIGNAL 1 24  
 FT CHAIN 25 816  
 FT DOMAIN 25 721  
 FT TRANSMEM 722 742  
 FT DOMAIN 743 816  
 FT DOMAIN 465 630  
 FT REPEAT 465 512  
 FT REPEAT 513 554  
 FT REPEAT 555 593  
 FT REPEAT 594 630  
 FT CARBOHYD 347 347  
 FT CARBOHYD 460 460  
 FT CARBOHYD 479 479  
 FT CARBOHYD 505 505  
 FT CARBOHYD 586 586  
 FT CARBOHYD 654 654  
 FT CARBOHYD 705 705  
 FT VARSPLIC 760 816  
 FT  
 FT CONFLICT 96  
 FT CONFLICT 259 259  
 FT CONFLICT 595 595  
 SQ SEQUENCE 816 AA; 90034 MW; 8ED736C097EBA855 CRC64;

Query Match 90.7%; Score 49; DB 11; Length 816;  
 Best Local Similarity 77.8%; Pred. No. 0.56;  
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 CDPGYIGSR 9  
 Db 489 CDPGYIGTR 497  
 |||||:|

RESULT 2  
 O9GK49  
 ID O9GK49 PRELIMINARY; PRT; 791 AA.  
 AC O9GK49;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE INTEGRIN BETA-5 SUBUNIT (FRAGMENT)  
 OS Bos taurus (Bovine)  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;  
 OC Bovidae; Bovinae; Bos.  
 OC NCBI\_TaxID=9913;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC TISSUE=MAMMARY GLAND;  
 RA Andersen M.H.; Rasmussen J.T., Berglund L., Petersen T.E.;  
 RT "Bovine beta-5 integrin subunit (fragment).";  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF317198; AAG38594.1; -;  
 FT NON\_TER 1  
 SQ SEQUENCE 791 AA; 87144 MW; 3FB45E62374169BC CRC64;

Query Match 81.5%; Score 44; DB 6; Length 791;  
 Best Local Similarity 66.7%; Pred. No. 4.8;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 CDPGYIGSR 9  
 Db 480 CNPGYIGTR 488  
 |||||:|

RESULT 3  
 O9Y0A4  
 ID O9Y0A4 PRELIMINARY; PRT; 446 AA.  
 AC O9Y0A4;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DE RAB GDP-DISSOCIATION INHIBITOR  
 GN RAB-GDI.  
 OS Branchiostoma floridae (Florida lancelet) (Amphioxus).  
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;  
 OC Branchiostoma.  
 OC NCBI\_TaxID=7739;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC Sedlacek Z., Shmied S.M., Muenstermann E., Poustka A.;  
 RT "The amphioxus RAB gdp-dissociation inhibitor (GDI) gene is neural -  
 RT specific: Implications for the evolution of function of chordate RAB  
 RT GDI genes.";  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: Y18521; CAB46230.1; -;  
 DR HSSP; P21856; 1GND.  
 DR InterPro: IPR002005; -;  
 DR InterPro: IPR002673; -;  
 DR InterPro: IPR003006; -;  
 DR Pfam: PF00996; GDI; 1.  
 DR PRINTS; PRO0891; RABGDIREP.  
 DR ProDom: PD010314; -; 1.  
 DR PROSITE; PS00290; IG\_MHC; UNKNOWN\_1.  
 SQ SEQUENCE 446 AA; 49759 MW; 69256AFFC0188D3A CRC64;

Query Match 77.8%; Score 42; DB 5; Length 446;  
 Best Local Similarity 66.7%; Pred. No. 6.3;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 CDPGYIGSR 9  
 Db 284 CDPGYAGDK 292  
 |||||:

RESULT 4  
 O9H3S8  
 ID O9H3S8 PRELIMINARY; PRT; 502 AA.  
 AC O9H3S8;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE NOTCH4 (FRAGMENT)  
 GN NOTCH4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC Miyagawa T.;  
 RA "Notch4 variant.";  
 RT Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.



QY 1 CDPGYIGSR 9  
I ||| |||  
Db 463 CPGYIGSR 471

RESULT 8  
Q20535 PRELIMINARY; PRT; 1827 AA.  
ID Q20535;  
AC Q20535;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE SIMILARITY TO EGF-TYPE REPEATS.  
GN F47C12.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
RA Parsons N., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,  
RA Snaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstein L., Wilkinson-Sproat J., Wohlman P.,  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RL elegans."  
RN Nature 368:32-38(1994).  
RP [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX Murray J., Wohlmann P., Sansone J.;  
RA Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX Waterston R.;  
RA Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX Waterston R.;  
RA Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX HSSP; U61946; AAC24388.1; -;  
DR EMBL; U61946; AAC24388.1; -;  
DR HSSP; P10998; 1VVD  
DR InterPro; IPR000152; -;  
DR InterPro; IPR000421; -;  
DR InterPro; IPR000436; -;  
DR InterPro; IPR000561; -;  
DR InterPro; IPR000859; -;  
DR InterPro; IPR001881; -;  
DR InterPro; IPR003410; -;  
DR Pfam; PF00008; EGF\_6;  
DR Pfam; PF00084; sushi; 7.  
DR Pfam; PF00431; CUB; 1.  
DR Pfam; PF00754; CUB; 1.  
DR Pfam; PF02494; FS\_F8\_type\_C; 1.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 2.  
DR PROSITE; PS01180; CUB; 1.  
DR PROSITE; PS00022; EGF\_1; UNKNOWN\_4.  
DR PROSITE; PS01186; EGF\_2; 4.  
DR PROSITE; PS01187; EGF\_CA; 2.  
DR SMART; SM00001; EGF\_like; 1.  
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
SQ SEQUENCE 1827 AA; 197946 MW; F966D515B58967AI CRC64;

Query Match 77.8%; Score 42; DB 5; Length 1827;  
Best Local Similarity 66.7%; Pred. No. 28;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CDPGYIGSR 9  
I ||| |||  
Db 1695 CKPGYIGSR 1703  
RESULT 9  
Q99940 PRELIMINARY; PRT; 1999 AA.  
ID Q99940;  
AC Q99940;  
DT 01-MAY-1997 (TREMBlrel. 03, Created)  
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE NOTCH4.  
GN NOTCH4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC Li L., Huang G., Banta A., Deng Y., Chen L., Pham Q., Rowen L.,  
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
RN EMBL; U99335; AAC63097.1; -;  
DR HSSP; P08709; 1BF9.  
DR InterPro; IPR000152; -;  
DR InterPro; IPR000561; -;  
DR InterPro; IPR000742; -;  
DR InterPro; IPR000800; -;  
DR InterPro; IPR001438; -;  
DR InterPro; IPR001881; -;  
DR InterPro; IPR002110; -;  
DR Pfam; PF00008; EGF; 27.  
DR Pfam; PF00023; ank; 6.  
DR Pfam; PF00066; notch; 2.  
DR PRINTS; PR00010; EGF\_BLOOD.  
DR PROSITE; PS50088; ANK\_REPEAT; 5.  
DR PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 11.  
DR PROSITE; PS00022; EGF\_1; UNKNOWN\_28.  
DR PROSITE; PS01186; EGF\_2; 21.  
DR PROSITE; PS01187; EGF\_CA; 9.  
DR SMART; SM00001; EGF\_like; 1.  
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
SQ SEQUENCE 1999 AA; 209134 MW; A978027EDA4E92A CRC64;

Query Match 77.8%; Score 42; DB 4; Length 1999;  
Best Local Similarity 77.8%; Pred. No. 31;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CDPGYIGSR 9  
I ||| |||  
Db 462 CPGYIGSR 470  
RESULT 10  
Q00306 PRELIMINARY; PRT; 2003 AA.  
ID Q00306;  
AC Q00306;  
DT 01-JUL-1997 (TREMBlrel. 04, Created)  
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE NOTCH4.  
GN HNOTCH4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI\_TaxID=9606;

[1]

SEQUENCE FROM N.A.

TISSUE=BONE MARROW, AND HEART;

Li L., Huang G., Banta A., Yu D., Rowen L., Hood L.;

Submitted (Mar-1997) to the EMBL/GenBank/DBJ databases.

EMBL; U95299; AAC32288.1; -

HSSP; P08709; 1BF9.

InterPro: IPR000152; -

InterPro: IPR000561; -

InterPro: IPR000742; -

InterPro: IPR000800; -

InterPro: IPR001438; -

InterPro: IPR001881; -

InterPro: IPR002110; -

Pfam; PF00008; EGF; 27.

Pfam; PF00023; ank; 6.

Pfam; PF00066; notch; 2.

PRINTS; PS0010; EGFELOOD.

PROSITE; PS00088; ANK\_REPEAT; 5.

PROSITE; PS0297; ANK\_REPEAT\_REGION; 1.

PROSITE; PS0010; ASX\_HYDROXYL; 11.

PROSITE; PS0022; EGF\_1; UNKNOWN\_28.

PROSITE; PS01186; EGF\_2; 21.

PROSITE; PS01187; EGF\_CA; 9.

SMART; SM00248; ANK; 1.

Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.

SEQUENCE 2003 AA; 209620 MW; BED10283343AOC14 CRC64;

Query Match 77.8%; Score 42; DB 4; Length 2003;

Best Local Similarity 77.8%; Pred. NO. 31;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9 77.8%; Score 42; DB 4; Length 2003;

DB 463 CPGYTGSR 471

RESULT 11

Q25253 PRELIMINARY; PRT: 2653 AA.

AC Q25253;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)

DT NOTCH HOMOLOG SCALLOPED WINGS (SCL).

DE ScL.

GN Lucilia cuprina (Greenbottle fly) (Australian sheep blowfly).

OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Cecidoidea; Calliphoridae; Lucilia.

OC NCBI\_TaxID=7375;

OX [1]

RN SEQUENCE FROM N.A.

RP SEQUENCE OF 39-265 FROM N.A.

RC STRAIN=SS SEEKING;

RX MEDLINE=96400928; PubMed=8807304;

RA Davies A.G., Game A.Y., Chen Z., Williams T.J., Goodall S., Yen J.L.,

RA McKenzie J.A., Batterham P.;

RA "scalloped wings is the Lucilia cuprina Notch homologue and a

RT candidate for the modifier of fitness and asymmetry of diazihon

RT resistance.;

RT Genetics 143:1321-1337(1996).

RL [2]

RN SEQUENCE OF 39-265 FROM N.A.

RP STRAIN=SS SEEKING;

RC Chen Z., Newsome T., McKenzie J.A., Batterham P.;

RA Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.

RL [3]

RN SEQUENCE OF 39-265 FROM N.A.

RP STRAIN=SS SEEKING;

RC Chen Z., McKenzie J.A., Batterham P.;

RA Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RL Waterston R.;  
RM Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U58746; AB00626.1; -;  
DR InterPro: IPR000561; -;  
DR Pfam: PF00008; EGF\_6;  
DR PROSITE: PS00022; EGF\_1; UNKNOWN\_7.  
DR PROSITE: PS01186; EGF\_2; 6.  
DR SMART: SM00181; EGF; 1.  
KW EGF-like domain; Glycoprotein; Hypothetical protein.  
FT DOMAIN 90 98  
SQ SEQUENCE 372 AA; 39085 MW; DB36AB251EEB6884 CRC64;

Query Match 75.9%; Score 41; DB 5; Length 372;  
Best Local Similarity 77.8%; Pred. No. 8;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CDPGYGSR 9  
DB 272 CAPGYGSR 280  
| | | | |

RESULT 13  
ID Q9VOA9 PRELIMINARY; PRT; 838 AA.  
AC Q9VOA9;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DE CG15388 PROTEIN.  
GN CG15388.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
RN [1] TaxID=7227;  
RP SEQUENCE FROM N.A.  
RC STRAIN-BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Ananadides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Wan K.H., Doyle C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Miklos G.L.G.,  
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalaal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RT Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of Drosophila melanogaster."  
RL Science 287:2185-2195(2000).  
DR EMBL; AE003583; AAF51268.1; -;  
DR HSP; P00740; IEDM.  
DR Flybase; FBgn0031414; CG15388.  
DR InterPro: IPR000152; -;  
DR InterPro: IPR000561; -;  
DR InterPro: IPR000742; -;  
DR InterPro: IPR001438; -;  
DR InterPro: IPR001881; -;  
DR InterPro: IPR002965; -;  
DR Pfam: PF00008; EGF; 5.  
DR PRINTS; PR00010; EGFBLD.  
DR PRINTS; PR01217; PRICHEXTENS.  
DR PROSITE; PS00010; ASX\_HYDROXYL; UNKNOWN\_3.  
DR PROSITE; PS00022; EGF\_1; 5.  
DR PROSITE; PS01186; EGF\_2; 3.  
DR PROSITE; PS01187; EGF\_CA; 3.  
DR SMART; SM00179; EGF\_CA; 1.  
KW Calcium-binding; EGF-like domain; Glycoprotein; Repeat.  
SQ SEQUENCE 838 AA; 87786 MW; A654CDBE4FDF4280 CRC64;

Query Match 75.9%; Score 41; DB 5; Length 838;  
Best Local Similarity 75.0%; Pred. No. 19;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CDPGYGS 8  
DB 39 CQPGYGS 46  
| | | | |

RESULT 14  
Q9TVQ2  
ID Q9TVQ2 PRELIMINARY; PRT; 1664 AA.  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DE Y64G10A.7 PROTEIN.  
GN Y64G10A.7  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1] SEQUENCE FROM N.A.  
RP SEQUENCE FROM N.A.  
RL Mortimore B.J.;  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2] MEDLINE=99069613; PubMed=9851915;  
RX SEQUENCE FROM N.A.  
RT none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
RT investigating biology."  
RL Science 282:2012-2018(1998).  
RN [3] SEQUENCE FROM N.A.  
RP Ainscough R.;  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL117206; CAB60454.1; -;  
DR EMBL; AL110498; CAB60454.1; JOINED.  
DR EMBL; AL117206; CAB57911.1; -;  
DR HSP; P00736; IAPQ.  
DR InterPro: IPR000152; -;  
DR InterPro: IPR000561; -;  
DR InterPro: IPR001881; -;



DR InterPro; IPR002049; -.  
DR Pfam; PF00008; EGF; 27.  
DR PRINTS; PR00011; EGF\_LAMININ.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 4.  
DR PROSITE; PS00022; EGF\_1; UNKNOWN\_22.  
DR PROSITE; PS01186; EGF\_2; 24.  
DR PROSITE; PS01187; EGF\_CA; 3.  
DR SMART; SM00179; EGF\_CA; 1.  
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.  
SQ SEQUENCE 1664 AA; 179279 MW; A69F093B4C705832 CRC64;

Query Match 75.9%; Score 41; DB 5; Length 1664;  
Best Local Similarity 66.7%; Pred. No. 40;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9  
| | | | | | | | | |  
Db 1360 CPPGYIGTK 1368

RESULT 15  
Q9GPA5 PRELIMINARY; PRT; 2524 AA.  
AC Q9GPA5;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE PUTATIVE NOTCH RECEPTOR PROTEIN.  
GN NOTCH.  
OS Branchiostoma floridae (Florida lancelet) (Amphioxus).  
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;  
OC Branchiostoma.  
OX NCBI\_TaxID=7739;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-WHOLE LARVAE;  
RA Holland L.Z., Burgtorf C., Holland N.D., Lehrach H., Tamme R.,  
RA Abi-Rached L., Pontarotti P., Lardelli M.;  
RT "Cloning and developmental expression of the amphioxus homologue of  
RT Notch (AmphiNotch): evolutionary conservation of multiple expression  
RT domains in amphioxus and vertebrates."  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Y12539; CAC19873.1; -.  
KW Receptor.  
SQ SEQUENCE 2524 AA; 270969 MW; C2CA57E306D23BC9 CRC64;

Query Match 75.9%; Score 41; DB 5; Length 2524;  
Best Local Similarity 66.7%; Pred. No. 62;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CDPGYIGSR 9  
| | | | | | | | | |  
Db 1236 CRPGYVQGR 1244

**THIS PAGE BLANK (USPTO)**